EXPERT SYSTEMS AND MULTIVARIATE ANALYSIS IN CLINICAL CHEMISTRY

Expertsystemen en multivariate analyse
in de klinische chemie

PROEFSCHRIFT

Ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof. Dr C.J. Rijnvos
en volgens besluit van het College van Dekanen.
De openbare verdediging zal plaatsvinden op
woensdag 11 november 1992 om 13.45 uur

door

Raymond Wilfred Wulkan
geboren te Pretoria (Zuid-Afrika)

ICG Printing
Dordrecht

PROMOTIE-COMMISSIE

Promotoren: Prof. Dr B. Leijnse
Prof. Dr E.S. Gelsema

Overige leden: Prof. Dr P.F. de Vries Robbé
Dr J. Lindemans

De uitgave van dit proefschrift is financieel ondersteund door de E.C.Noyons Stichting (Stichting ter Bevordering van de Klinische Chemie in Nederland).

Cette absorption d'oxygène aux globules du sang n'exige d'ailleurs aucune force vitale de nature mystérieuse. Elle dépend de conditions purement chimiques.

Leçons sur les phénomènes de la vie (1879) Claude Bernard

Aan mijn moeder
Ter nagedachtenis aan mijn vader

CONTENTS

Chapter 3
LITHOS and CALCULI, two computer programs for the evaluation of X-ray diffractograms of urinary calculi

Chapter 4
Interpretation of acid-base measurements

Chapter 5
Multivariate reference regions

Chapter 1

INTRODUCTION

1.1 Developments in clinical chemistry

Clinical chemistry is the science occupied with the analysis of chemical constituents in body fluids and with conversion of the resulting data into information with a higher abstraction level. The developments in clinical chemistry in the last decades have occurred in rapid succession. Improvements in technology permitted the introduction of autoanalyzers, thereby allowing an immense increase in capacity of the clinical laboratory. With the expansion of biochemical research, a growing interest in the diagnostic possibilities ensued, resulting in a steady increase in daily workload. From the clinical chemistry laboratory of our hospital with 1000 beds, every day approximately 8000 analytical results are transmitted from the laboratory to the clinic. Developments in computer technology enabled a more practical handling of the increased flow of data. These data can be automatically transmitted from the laboratory to the clinic, and they can be stored and accumulated to large sets for later inspection. These data provide information when interpreted against reference ranges, taking their interdependency into account, and when interpreted in relation to the clinical picture. Attempts have also been made toward the interpretation of laboratory data leading to laboratory diagnoses.

1.2 Reference ranges

Laboratory data are interpreted primarily against a range of values that are found in healthy control subjects (1). Generally, a 95% confidence interval obtained from measurements in material from such control persons is used as the reference range. This interval is often calculated assuming a Gaussian distribution.
Some criticism may be given concerning this approach. Not all analytes follow a Gaussian distribution curve (2,3). Furthermore, there are many factors that may influence the reference range. For many analytes there is a variation in reference range with sex and age. The concentration of some analytes depends on the posture of the body, *e.g.* the serum total protein concentration is known to decrease by 12% if the patient is allowed to lie down for half an hour (4). Some analytes follow a circadian rhythm. By consequence, the sampling time is important in the interpretation of the result. The concentration of some analytes is influenced by diet.

In the light of these comments, the determination of reference intervals from the hospital population seems advantageous. In this case, the important assumption is made that the majority of laboratory results is not pathological, even though it has been obtained from a patient population subject to pathology. Although this presupposition has been adopted by many without much comment, it seems important to find the reason for this paradoxical situation. The central part of a sufficiently large number of patient results often follows a Gaussian distribution. This central Gaussian part is not likely to be characterized by a very pathological mean value. There are several arguments to support this postulate. Experience learns that usually a considerable part of all measurements is within the normal range. Furthermore, if only the central Gaussian part of the patient population is selected for the calculation of the reference range, there is no interference from extreme pathological values. The less extreme values at both sides of the central part will influence the calculated mean and standard deviation. If these values at both sides of the central part are asymmetrically distributed, they will not pass the test for Gaussian behaviour and hence be excluded from calculation (*i.e.* no influence). If they are symmetrically distributed, they will annihilate each others influence on the mean value, but they will still have an upward effect on the calculated standard deviation. The shift of the mean of the central part away from the normal range, will be discussed later (ch. 5). A prerequisite to the approach mentioned above is that the size of the Gaussian fraction should be sufficiently large to enable its mathematical separation from the pathological fraction. By this approach, the continuous monitoring of the reference values obtained from the hospital population is obligatory. With a drift in age distribution for instance, it seems likely that the reference range should be adjusted accordingly (5). However, this is not a decisive argument against the use of patient-based reference ranges, because in the conventional approach (healthy reference subjects) a re-evaluation with an adjusted age distribution would be just as necessary.

1.3 Multivariate analysis

In general, the calculation and application of reference values has been done in an univariate manner. The result of a single analysis is interpreted against a 'silver' standard, being the reference range of the laboratory where the analysis is done.
When the results of two types of analysis A and B, obtained with the same sample, are taken into account simultaneously, cases occur for which the results fall within the reference range for analysis A, but outside the reference range for analysis B, and *vice versa*. In this way, the fraction of normal individuals that have both analyses within their reference range is smaller than the fraction that has only one within range. When more types of analysis are considered together, the fraction of normal individuals having all values within range decreases rapidly. It can be calculated, assuming 95% reference ranges, that only 77% of normal individuals will have all their values within range when

five types of uncorrelated analyses are considered together.
When two correlated analyses are considered together, the reference range becomes an ellipse, of which the shape and orientation is dependent on the correlation coefficient (fig. 1.1).

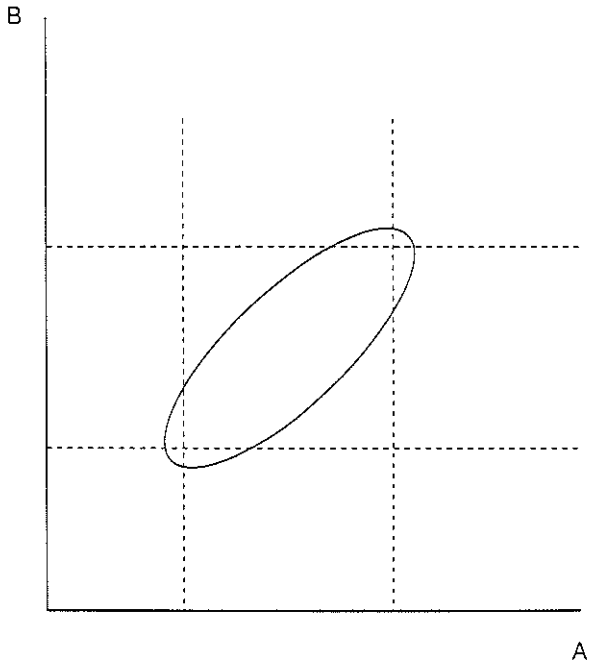


Figure 1.1 *Univariate and multivariate reference ranges.*

For those analyses that are strongly correlated, the ellipse becomes shallow and will eventually become a straight line. In this hypothetical situation a measurement of A means that B can be calculated from the relation. In the complementary situation with a very weak correlation between A and B, the shape of the ellipse approaches a circle (provided that A and B have been normalized by division by their respective standard deviation). In this situation, the use of multivariate analysis is of little value because the number of patient results that are outside the circle will not be very different from the number being outside the square formed by both univariate reference ranges. Put in another way, a multivariate approach is advantageous with parameters having a tight correlation, because there is a larger area where abnormalities may be signalled. Although the probability to find such a result is likely to decrease with increasing correlation

coefficient, the information contained in such a combination has either pathological or analytical significance. In chapter 5 we will further expand on this matter.

1.4 Information enhancement in clinical chemistry

Generally, laboratory tests are requested for either diagnosis, monitoring of the patients functions or therapy. The results are objective data that reflect the current condition of the patient. These data are subject to some amount of variation from the pre-analytical phase and to a known measurement error.
In contrast, clinical information is of a much more subjective nature, with an exception for information that can be obtained by measurements (*e.g.* pulse rate or body temperature). Because the signs and symptoms of disease are the result of complex multifactorial processes, the application of clear-cut rules laid down in algebraic *formulae* in this field is an impossibility.
Many clinical chemistry variates are subject to physical and chemical laws, *e.g.* the principles of electroneutrality (reflected in the anion gap, refs.6,7) and iso-osmolality (reflected in the osmolality gap, ref.8), the pH-dependent binding of calcium to the plasma proteins, and the Henderson-Hasselbalch equation governing the acid-base equilibrium. Several of these formalizations rely on the assumption of an equilibrium situation and a system free from interfering compounds. In systems with compartmentalization, progressive complexity, and dynamic behaviour, the application of these principles becomes increasingly difficult.
Nevertheless, clinical chemistry analyses provide a wealth of basal, hard-core data, that can be combined to yield information of a higher abstraction level. This transformation can be done by expert systems (see section 1.6) incorporating a mixture of pattern recognition, numerical formula evaluation, and reasoning based on physiological principles. For instance, in analyses of creatinine and urea in 1488 sera, only 4 out of the 9 (3^2) possible patterns occurred consistently when both analytes were classified as low, normal, or high. In the same study, 7695 sera were evaluated for sodium, potassium, chloride, and bicarbonate, and of the 81 (3^4) possible patterns, only 41 were actually observed (9). Unusual patterns may serve to detect laboratory error, or to attend the physician to an exceptional situation. Multivariate analysis of large data sets can provide statistical information that can be incorporated in expert systems. For instance, the probability may be calculated that a certain pattern is associated with a certain disease.
The enhanced information produced from clinical chemistry analyses must serve as a functional support to the physician in the diagnostic process or in the monitoring of the patients condition or therapy.

1.5 The need for interpretative reporting

In one study, 71% of the physicians questioned believed they did not require computer assistance for acid-base analysis. However, when actually tested, they were able to correctly identify only 39% of acid-base problems (10). In another study, the correct response rates of 21 physicians to single, double, and triple acid-base disorders were 86%, 49%, and 17%, respectively (11). In a study on the diagnosis of anemia, abundant laboratory data were found (*i.e.* normal vitamin B_{12}/folate levels in iron deficiency anemia) that would be considered redundant retrospectively (12).
In several other studies, a number of cases occurred in which physicians had to admit that the computer interpretation was correct and they adjusted their diagnosis accordingly (13,14).

1.6 Expert systems

1.6.1 Definition

According to an often used definition, an expert system is a computer program that can give advice in a well-defined area of expertise and should be able to explain its line of reasoning.

1.6.2 History

Because the development of expert systems has its roots in artificial intelligence, the development of the first expert systems originated at large research institutes like the Stanford University, the Massachusetts Institute of Technology and the Carnegie-Mellon University. One of the first prototypes was DENDRAL, an expert system designed to solve problems in mass spectrometry (15). The first system in the field of medicine was MYCIN, designed at Stanford University for the diagnosis and therapy of infectious disorders (16). An expert system is characterized by separation of the expert knowledge from the surrounding program. When this was done with MYCIN, the surrounding program remained as an empty shell and was called EMYCIN (Essential or Empty MYCIN). This part of MYCIN can be used to build expert systems in other fields of knowledge.
The early large expert systems like INTERNIST-1 were run on mainframes (17). Nowadays there is a trend towards smaller computers: workstations, personal computers and even an occasional pocket calculator.
A wide scale in magnitude of knowledge domains may be found: large systems like INTERNIST-1 (internal medicine, ref.17), systems of intermediate size like HEPAR (hepatology, ref.18), smaller systems like BCDE2 (microcytic anemia; ch.6) and systems with a very small domain like erythrocyte enzymes or isoenzymes of cholinesterase (section 1.9).

1.6.3 Languages and shells

Many computer languages have been used to develop expert systems; the eldest are LISP (List Processing Language) and PROLOG (Programming Logic). A great variety of LISP dialects exists. The major part of this thesis deals with expert systems in a LISP environment.
Numerous empty shells have been developed, of which a few examples are given: EMYCIN (LISP, Stanford University), Personal Consultant Plus (LISP, Texas Instruments), EXPERT (FORTRAN, Rutgers University, ref.19), SMR (Israel Institute of Technology, ref.20), KEE (LISP, Intellicorp), Acquaint (LISP, LithP Systems), DELFI (TurboPascal, Delft University of Technology, ref.21), Goldworks (LISP, Bolesian Systems Europe), Pro.M.D. (PROLOG, ref.22), Nexpert Object (C) etc. These shells were developed to enable the building of expert systems without knowledge of high level computer languages like LISP or PROLOG.

1.6.4 Structure

Traditionally, an expert system consists of a part containing the expert knowledge (the knowledge base) and a part driving the process of reasoning (the inference engine). The knowledge is represented in rules with a PREMISE and an ACTION part similar to IF..THEN structures. The process of reasoning can be either data driven, goal driven, or a combination of both. In the first case new facts are deduced from input data (forward chaining). In the second case possible conclusions are verified by retrieval of data leading to acceptation or rejection (backward chaining). A combination of both mechanisms is a more powerful approach and more resemblant to human problem solving.
The rules in an expert system are grouped into contexts, which are rulegroups that functionally belong together in dealing with a specific part of the problem. This structure is helpful in keeping an overview when the expert system development becomes complicated. Rules are used to conclude facts, to evoke programs (internally in LISP, or externally in *e.g.* PASCAL, FORTRAN, or C), or to guide the inference process. The latter rules are called meta-rules because they contain knowledge about the other rules, and thus represent a higher level of 'intelligence'. They can be used for example to compel the system to 'change its mind' about what part of the knowledge has to be consulted, or to 'forget' some tasks it had planned to do, in cases where these are no longer relevant.
In the newer shells, more emphasis has been laid on an effective communication with databases. Thus, a situation of complete automation can be created where the time-inefficient question-and-answer interplay with the user is avoided and data can be retrieved from a database, which is also less prone to human error. In the reverse situation, conclusions made by an expert system may be stored in the database.

1.6.5 Function

The definition of an expert system mentioned above is divided between two functions: one of expert judgement and one of expert teaching. In our opinion these functions should not be combined in the same system, since the expert system as a tutor carries a large amount of overweight which is not needed in a consultancy situation.
The ability of expert systems to explain their line of reasoning is largely overestimated. Generally, the user may ask, by means of keyboard input, *why* the system requires additional data, or *how* the system arrived at the current intermediate conclusion. The system may reply that the data was needed in order to trace the value of a certain parameter or alternatively, that it reached its conclusion by applying the action part of the current rule (which is then given). Of course this is far from a situation of teaching the art to a novice. In our opinion, systems for tutorial situations need to be developed separately, because the kind of knowledge needed for tutorial and explanatory situations is different from that of giving expert judgement. For completeness, it must be mentioned that these explanatory functions can be quite useful in the development phase of an expert system.
Expert systems can be used in several ways. One is a function as a 'watchdog' to select the cases that deserve the detailed attention of a human expert. Here, the bulk of relatively simple cases are treated without human intervention and the system largely acts as a time-saver for the real expert. Such systems have a rather broad but shallow knowledge base.
Another function is that of an expert only for complicated cases. These systems typically have a large knowledge base limited to a small area of expertise. In this case the conclusions generated by such a system should be accepted only after being reviewed by a human expert. Usually, this expert will not be impressed by the advice of his binary colleague because he would have given the same advice. There may arise differences of opinion between the program and the expert if the latter was not involved in the development of the system. Disagreement between human experts has been documented repeatedly (23-25). Especially for this type of expert system the development phase never ends because it has to be continuously refined according to the state of the art.

1.7 Problems in developing expert systems

Several difficulties may arise in the development of an expert system. Depending on the domain of knowledge, the choice of the appropriate hard- and software is a major issue. We experienced that in the long run the 640 kb internal memory of our Olivetti M24 SP personal computer became a limitation for the developing program. A survey of the literature shows that the larger systems have all been built on mainframes or workstations. On the other hand, it has always been stressed that the choice of too large a knowledge

domain should be avoided. Indeed, many systems have been developed on personal computers in conventional programming languages.
Another problem is the collection of information from different locations (apparatus) that is not produced simultaneously. For simple situations this may be done by human intervention. Alternatively, information may be obtained from the hospital information system or from a laboratory database. Apart from this timing problem, the situation becomes more complex when clinical information is to be part of the input data. Although age, sex, and in some cases medication may be obtained from the hospital information system, simple data like length, weight or body temperature will dictate human intervention, and thereby inefficiency. Such data can only be obtained through co-operation of non-laboratory staff. We therefore advocate the development of systems that are limited to laboratory matters. These could well make use of an (internal) working hypothesis but they should never attempt to make a *clinical* diagnosis because the necessary information is lacking. We expect that this approach will prove helpful in attaining acceptance for these systems.
Before an analysis of the knowledge domain is made, it has to be decided whether the expert and the knowledge engineer will work as a team or whether the expert is to implement the knowledge himself. Some groups have developed special interviewing techniques because it was difficult for the experts to explain all details of their reasoning. In the development of LITHOS (ch. 2), we have found it rewarding to take a short course in expert system development and do the job ourselves, working in a team of two persons.
Another well-known problem in expert system development concerns the performance at the border of the knowledge domain, *i.e.* can the system decide whether the problem is (partially) beyond its capability. Although this is not likely to present a real problem in clinical chemistry, this difficulty can be dealt with by meta-rules. Furthermore, it is important to keep in mind that the transfer of expert systems between hospitals is not always possible. Another 'school' of decision making, or other analytical methods may be present and systems that depend on population studies may show failure.

1.8 Legislation

Using an expert system that has to contribute information for clinical decisions means dealing with the question of responsibility. Who is to be held responsible if the system contributes with faulty information that could result in a decision that ultimately proves fatal to the patient? Is it the over-confident expert who provided the knowledge to be incorporated in the program; is it the knowledge engineer who failed to detect that the program was not flawless; or is it the user of the program, be it a clinical biochemist or a physician, who omitted to convince himself of the impeccability of the system? The answer to these questions depends on the setting in which it occurs. Although literature

about this subject is rather scarce, some considerations about the German situation have been published by Hoffmann (26). It seems logical to accept that the responsibility should lie with those who apply such systems to provide information for the clinic. The situation can be compared to the purchase of a newly developed analytical instrument, for which flawless behaviour is claimed by the manufacturer. The buyer should convince himself of the reliability of such an instrument when it is used in daily routine. In the same way that he is responsible for the analytical quality of its results, he should also stand for the quality of the information that is provided by expert systems.

1.9 Applications in clinical chemistry

An overview of applications is given in Table 1.1.

Table 1.1 *Overview of interpretative systems for clinical chemistry.*

Subject	Literature reference
Acid-base disorders	See chapter 4, Table 4.4
Electrolyte disorders	27,28
Anemia	See chapter 6, Table 6.9
Myelodysplastic syndromes	29
Leukemia	30,31
Erythrocyte enzymes	32
Transfusion reactions	33
Red-cell antibody identification	34
Platelet request evaluation	35,36
Bleeding disorders	37
Complement factors	38
Calcium metabolism	39,40
Calcium/phosphorus metabolism	41
Alkaline phosphatase isoenzyme patterns	42
Cholinesterase isoenzyme patterns	43
Hepatic enzymes	44
Serum and CSF protein electrophoresis	45-47
Aspartate aminotransferase elevations	48
Urinalysis	49
Lipoprotein disorders	14,50-52
Thyroid disorders	53,54
Renal stone analysis	55-58
Instrument troubleshooting	59,60
Validation of analytical results	61

1.10 Future developments

The areas where expert systems may find a fruitful application begin to delineate. There have been encouraging developments in the troubleshooting of analyzers (59,60). Another field of interest is accelerated laboratory investigation (62,63). Following this approach, analyzers may be equipped with expert systems that generate the next logical test from the current test results. When fiatted by the physician, this next test may be carried out using the same sample. This saves time of both physician and patient, it is less discomforting to the patient because only one rather than two samples are needed, and it enables a systematic approach in the diagnostic work-up.
In the future, expert systems coupled to the hospital information system may retrieve information about the medication and decide if an aberrant test result was caused by drug interference. Large databases in this field are commercially available nowadays. This application is in fact part of the plausibility control, which could become another field of application for expert systems. At the same time, information could be generated by multivariate analysis of suitable combinations of parameters (ch. 5). The same technique could be used to determine which tests should be rerun because of questionable plausibility.
In our view, the incorporation of the time factor could become rewarding as well. In this way, a dynamic instead of quasi-static control of a patient is possible. This approach calls for rapid retrieval of historical information, either from a laboratory database or from the hospital information system.

1.11 Scope

The object of this thesis is to explore how expert systems may be helpful in information enhancement for clinical chemistry. Several prototypes have been built in order to investigate the advantages and disadvantages. A discussion of existing and new ideas with ample references has been added to create a birds eye view of the area, and to allow the interested reader a deeper dive into the subject. It is hoped that this thesis will act as a catalyst in reverting the existing scepticism into enthousiasm.
Chapter 2 is a description of LITHOS, an expert system for X-ray diffraction analysis of urinary calculi. A comparison of this system with CALCULI, a conventional computer program for the same purpose is given in chapter 3. Chapter 4 contains a discussion of acid-base nomenclature and a short description of CHEMPATH, an expert system for the diagnosis of acid-base and electrolyte disorders. In chapter 5, examples and a discussion are presented of multivariate analysis as a tool in clinical chemistry. Some explorations in developing an expert system for the laboratory diagnosis of anemia (HEMO) are discussed in chapter 6. Chapter 7 presents a general discussion of the application of expert systems in clinical chemistry, as well as the experiences with the various systems.

1.12 References

1. Solberg HE, Gräsbeck R
   Reference values. Adv Clin Chem 1989;27:1-79.

2. Hoeke JOO
   Aspecten van de 'gemiddelde van normalen'-methode in de klinische chemie. [Dissertation]. Rotterdam, The Netherlands: Erasmus Universiteit Rotterdam, 1976. 87 pp.

3. Naus AJM
   De behandeling van referentiewaarden in de klinische chemie uit analyseresultaten van een patientenpopulatie. [Dissertation]. Maastricht, The Netherlands: Rijksuniversiteit Limburg, 1982. 154 pp.

4. Bohme A
   Uber die Schwankungen der Serum-konzentration beim gesunden Menschen.
   Dt Arch Klin Med 1911;103:522-62.

5. Morgan DB
   The impact of ageing - present and future. Ann Clin Biochem 1983;20:257-61.

6. Emmett M, Narins RG
   Clinical use of the anion gap. Medicine 1977;56:38-54.

7. Gabow PA
   Disorders associated with an altered anion gap. Kidney Int 1985;27:472-83.

8. Dorwart WV, Chalmers L
   Comparison of methods for calculating serum osmolality from chemical concentrations, and the prognostic value of such calculations. Clin Chem 1975;21:190-4.

9. Lindberg DAB
   Collection, evaluation, and transmission of hospital laboratory data. Meth Inf Med 1988;11:20-30.

10. Hingston DM, Irwin RS, Pratter MR *et al.*
    A computerized interpretation of arterial pH and blood gas data: Do physicians need it? Resp Care 1982; 27:809-15.

11. Schreck DM, Zacharias D, Grunau CFV
    Diagnosis of complex acid-base disorders: physician performance versus the microcomputer. Ann Emerg Med 1986;15:164/97-170/103.

12. Blomberg DJ, Guth JL, Fattu JM, Patrick EA
    Evaluation of a new classification system for anemias using Consult Learning System. Comp Meth Progr Biomed 1986;22:119-25.

13. Bleich HL
    The computer as a consultant. New Engl J Med 1971;284:141-7.

14. Pincé H, Cobbaert C, van de Woestijne M, Lissens W, Willems JL
Computer aided phenotyping of dyslipoproteinemia. Int J Biomed Comput 1988;23:251-63.

15. Buchanan BG, Sutherland GL, Feigenbaum EA
Heuristic DENDRAL: a program for generating explanatory hypotheses in organic chemistry. In: Meltzer B, Mitchie D, eds. Machine Intelligence. Edinburgh: Edinburgh University Press. 1969;4:209-54.

16. Shortliffe EH
Computer-based medical consultations: MYCIN. New York: Elsevier Publishing Company (1976).

17. Miller RA, Pople HE, Myers JD
INTERNIST-1, an experimental computer-based diagnostic consultant for general internal medicine. New Engl J Med 1982;307:468-76.

18. Lucas PJF, Segaar RW, Janssens AR
HEPAR, an expert system for the diagnosis of disorders of the liver and the biliary tract. Liver 1989;9:266-75.

19. Weiss S, Kulikowski C
EXPERT: a system for developing consultation models. 4th Int Joint Conf Artif Intell, Tokyo, 1979:841-6.

20. Wiener F
SMR (simulating medical reasoning): an expert shell for non-AI experts. Comp Meth Prog Biomed 1988; 26:19-32.

21. de Swaan Arons H, Groen A, Hofland AG, Stoop JC
Building and consulting of expert systems with DELFI. Delft Progress Report 1984;9:63-85.

22. Pohl B, Trendelenburg C
Pro.M.D. - a diagnostic expert system shell for clinical chemistry test result interpretation. Meth Inf Med 1988;27:111-7.

23. Lanzola G, Stefanelli M, Barosi G, Magnani L
NEOANEMIA: a knowledge-based system emulating diagnostic reasoning. Comp Biomed Res 1990;23: 560-82.

24. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Gralnick HR, Sultan C
The morphological classification of acute lymphoblastic leukaemia: concordance among observers and clinical correlations. Br J Haematol 1984;47:553-61.

25. Ridderikhof C
Decision-making strategies in the general practice. [Dissertation]. Rotterdam, The Netherlands: Erasmus Universiteit Rotterdam 1986. 142 pp.

26. Hoffmann G
Arbeitsgruppe Geräte-evaluation: Interaktion zwischen mechanisierten Analysengeräten und wissensverarbeitenden Systemen. Mitt Dt Ges f Klin Chemie e.V. 1990;1:25-27.

27. Swezey CB, Jacobson C
Computer-based diagnostic reporting for serum electrolytes. Am J Clin Pathol 1980;74:812-9.

28. Leaning MS, Summerfield JA
Microcomputer-based management of fluid and electrolyte balance in hospitalized patients. In: Salamon R *et al.*, eds. MEDINFO86. Elsevier Science Publishers B.V. (North-Holland), 1986:138-41.

29. Imbert M, Priolet G, Dadi W, Sultan C
An expert system applied to the diagnosis of anemia with special reference to myelodysplastic syndromes. Blood Cells 1989;15:563-71.

30. Fox J, Myers CD, Greaves MF, Pegram S
Knowledge acquisition for expert systems: experience in leukemia diagnosis. Meth Inf Med 1985;24:65-72.

31. Alvey PL, Preston NJ, Greaves MF
High performance for expert systems: II. A system for leukemia diagnosis. Med Inf 1987;12:97-114.

32. Wiener F, de Verdier C, Groth T
The use of knowledge-based information systems for interpreting specialized clinical chemistry analyses-experience from erythrocyte enzymes and metabolites. Scand J Clin Lab Invest 1990;50:247-59.

33. Shifman MA, Vesilind GW
TREACT. An expert system consultation program to aid in the diagnosis of transfusion reactions. Transfusion 1988;28:253-6.

34. Smith JW, Svirbely JR, Evans CA *et al.*
RED: a red-cell antibody identification expert model. J Med Syst 1985;9:121-38.

35. Connely DP, Sielaff BH, Scott EP
ESPRE-Expert system for platelet request evaluation. Am J Clin Pathol 1990;94(suppl 1):19-24.

36. Sielaff BH, Scott EP, Connelly DP
Design and preliminary evaluation of an expert system for platelet request evaluation. Transfusion 1991;31:600-6.

37. Bennett JS, Goldman D
CLOT: a knowledge-based consultant for bleeding disorders. Memo HPP-80-7. Computer Science Department, Stanford University (1980).

38. Wiener F, Groth T, Nilsson U
A knowledge-based system for automatic interpretation of an analytical profile of complement factors. J Clin Lab Anal 1989;3:287-95.

39. Eberle F, Schmidt-Gayk H, Trendelenburg C, Katterman R
An expert system for the interpretation of the calcium metabolism. Automedica 1987;8:88.

40. Ledochowski M, Herold M, Dienstl F
A pocket computer program for differential diagnosis of tumor-induced hypercalcemia and primary hyperparathyroidism. Comp Meth Prog Biomed 1985;21:55.

41. Kinney EL, Wright RJ, Caldwell JW
The calcium-phosphorus metabolism expert system module: a program developed in the ALEX Smalltalk/V shell. J Med Syst 1989;13:49-53.

42. van Hoof VO, Pohl B, Verpooten GA, Lepoutre LG, de Broe ME
A knowledge based expert system for the clinical interpretation of alkaline phosphatase isoenzyme patterns. In: Köchli HP, Peheim E, eds. Proc 7th Int Conf 'Computing in clinical laboratories'. Lugano, Switzerland, 1989.

43. Loughlin JF, Tuckerman JF, Henderson AR
A BASIC program for serum cholinesterase phenotyping using a microcomputer. Ann Clin Biochem 1984;21:43-4.

44. Thiele H, Adam N, Irrgang B, Poegel K, Schulz H, Schwab A
Interpretation of laboratory test results in hepatobiliary diseases. In: Köchli HP, Peheim E, eds. Proc 7th Int Conf 'Computing in clinical laboratories'. Lugano, Switzerland, 1989.

45. Weiss SM, Kulikowski CA, Galen RS
Developing microprocessor based expert models for instrument interpretation. Proc 7th Int Joint Conf Art Int 1981:853-5.

46. Weiss SM, Kulikowski CA, Galen RS
Representing expertise in a computer program: the serum protein diagnostic program. J Clin Lab Automation 1983;3:383-7.

47. Przetak C, Wick M, Einhäupl K, Fateh-Moghadam A
Computer-assisted result presentation and interpretation of protein analyses in CSF and serum. In: Köchli HP, Peheim E, eds. Proc 7th Int Conf 'Computing in clinical laboratories'. Lugano, Switzerland, 1989.

48. Llaurado JG
EXPERT valuation of aspartate aminotransferase (AST) elevations: a forerunner of sequential laboratory testing. Int J Biomed Comput 1988;22:159-64.

49. Sweeley CC
Computer diagnoses disorders by analyzing urine. Medilab august/sept.1985:9-10.

50. Loughlin JF, Leung FY, Henderson AR
Classification of hyperlipoproteinaemias by computer interpretation. Ann Clin Biochem 1984;21:326-31.

51. van de Woestijne M, Cobbaert C, Lissens W, Willems JL
Fenotypering van dyslipoproteinemie: een expert systeem geschreven in PROLOG. Proc 7th Med Inf Congr. de Moor G, Stevens C, eds. Antwerpen, Belgium, 1987:115-20.

52. Trendelenburg C, Wieland H
Routine use of a clinical chemistry expert system with a knowledge base on disorders of lipoprotein metabolism. Mitt Dt Ges f Klin Chemie e.V. 1990;1:25-7.

53. Horn KA, Compton PJ, Lazarus L, Quinlan JR
An expert computer system for the interpretation of thyroid assays in a clinical laboratory. Aust Comp J 1985;17:7-11.

54. Degener FL, Santen R, Krüskemper HL
Ein Expertensystem in der Schilddrüsendiagnostik. Med Welt 1989;40:1139-44.

55. Wulkan RW, Zwang L, Liem TL, Blijenberg BG, Leijnse B
LITHOS, an expert system for interpretation of X-ray diffractograms of urinary calculi. J Clin Chem Clin Biochem 1987;25:719-22.

56. Wulkan RW, Leijnse B
Experience with expert systems in clinical chemistry. In: Kerkhof PLM, van Dieijen-Visser MP, eds. Laboratory data and patient care. New York and London: Plenum Press, 1988:117-24.

57. Blijenberg BG, Wulkan RW, Zwang L, Liem TL, Leijnse B
Computerunterstützte Auswertung der Röntgendiffraktometrie-Analysen. In: Hesse A, Claßen A, Röhle G, eds. Labordiagnostik bei Urolithiasis. Stuttgart: Wissenschaftliche Verlagsgesellschaft mbH, 1989:65-70.

58. Hesse A, Gergeleit M, Schüller P, Möller K, Brühl M
Computerunterstützte Auswertung der infrarotspektroskopischen Analyse. In: Hesse A, Claßen A, Röhle G, eds. Labordiagnostik bei Urolithiasis. Stuttgart: Wissenschaftliche Verlagsgesellschaft mbH, 1989:79-84.

59. v.d. Boogaard J
Bayer Technicon Instruments B.V., The Netherlands. A troubleshooting program for the H*1.

60. Woodbury WF
The development of a PC-based expert system to aid in the troubleshooting of clinical laboratory instrumentation. Lab Med 1989;20:176-82.

61. Valdiguié PM, Rogari E, Philippe P
VALAB: expert system for the validation of biochemical data. Clin Chem 1992;38:83-7.

62. Altshuler CH, Bareta J, Carafo AF, Carafo JR, Gibbon SL
PALI (Programmed Accelerated Laboratory Investigation). Crit Rev Clin Lab Sci 1972;3:379-402.

63. van Lente F, Castellani W, Chou D, Matzen RN, Galen RS
Application of the EXPERT consultation system to accelerated laboratory testing and interpretation. Clin Chem 1986;32:1719-25.

Chapter 2

LITHOS

AN EXPERT SYSTEM FOR THE INTERPRETATION OF X-RAY DIFFRACTOGRAM DATA

2.1 Introduction

Urinary calculi are found with a prevalence of approximately 4.7% in the Dutch population (1). The analysis of urinary stones provides the attending physician with data to install an appropriate treatment. The analysis of urinary stones is usually done by X-ray diffraction with a Debije-Scherrer-Hull camera. X-ray analysis and infrared spectroscopy are superior to chemical analysis, which is known to be hampered by a large fraction of false positive (mistakenly detected) and false negative (mistakenly missed) components (2). An extensive description of X-ray diffraction is beyond the scope of this thesis: the interested reader is referred to the publication of Dosch *et al.* (3).
An analysis with the Debije-Scherrer-Hull camera results in a pattern of several concentric dark rings on a photographic film (4). The radius of these rings is a measure of the crystal lattice distances, whereas the photographic density indicates the intensity of the reflected radiation, which is a measure of concentration. The interpretation of this pattern is done by specialists with many years of experience and yields the composition of the calculus, qualitatively as well as semi-quantitatively (mass percentages).
Treatment of patients with renal calculi is now routinely done by means of extracorporal shock wave lithotrypsy. This non-invasive technique has resulted in a considerable increase in requests for diffraction analysis (5). The ensueing need for a greater analytical capacity was satisfied by means of a diffractometer analyzer. The digitalized output from this instrument is transferred to a personal computer to be interpreted by an expert system.

2.2 Sample preparation

As a first step in the analysis of renal stone components, the complete calculus is ground in an agath mortar. The grinding should be carried on long enough to ensure that the particle size is sufficiently small. The monocrystalline silicon applicator disk is rubbed with Apiezon L grease (E. Merck, Darmstadt, Germany). The excess Apiezon is skimmed

off with a microscopic object glass. At least 4 mg of the powder is distributed on the disk. It is then gently tapped with its side on the mortar to remove the excess powder. This procedure results in a regularly distributed layer of the sample.

2.3 The diffractometer system

Analyses of urinary calculi are performed using a Compact X-ray Diffraction Analyzer System (PW 1840/10; Philips Nederland BV, Eindhoven, The Netherlands). This system consists of a high voltage generator (PW 1729), an X-ray tube with a copper anode, a goniometer driven by a stepping motor, a nickle filter, an X-ray detector, and a control unit. The latter is equipped with an integrator/interface board which permits remote control by means of a personal computer. The X-ray tube is operated at 40 kV/50 mA and emits $K\alpha_{1,2}$ radiation of 1.542 Å, and $K\beta$ radiation of 1.392 Å. The diffractometer output consists of diffraction angles (in steps of 0.01° through a range of 9.00–47.00°) and the corresponding intensities. The scan speed is a compromise between optimal peak resolution and minimal analysis time and is set to 0.05°/sec. The time constant (τ) of the apparatus is set to 0.5 sec. With these settings the intense peak of α-quartz at 26.72° is measured at 92% of its true value (τ/σ = 0.2, where σ is the standard deviation of the peak).
The diffraction angles (2Θ values, ref. 3) and intensities of the reflection are sent to an Olivetti M24 Personal Computer, equipped with a 20 Mb Winchester drive and a 5.25 inch floppy drive. The personal computer has an extended internal memory of 640 kb and an Intel 8087 coprocessor unit.

2.4 The diffractometer control program

A small demonstration program in GW BASIC for remote controlled operation was supplied by the manufacturer. The program was rewritten in TURBO BASIC and extended to enable data manipulation. The following features were added: storage of the raw data from the diffractometer, Savitzky-Golay 11-point filtering, noise threshold generation, curved baseline subtraction, calculation of the first derivative, peak selection, screen presentation, and storage of selected data in a format suitable to be used by the expert system.
In order to calculate the quantitative composition of the sample, the expert system LITHOS uses peak heights rather than peak surfaces. The integration of peaks in this automated setting is a nearly impossible task because of the many overlapping, irregularly shaped peaks, and the interference of small peaks from the sample matrix. As an illustration, figure 2.1 shows a typical diffractogram of Apatite, a component that is known for its microcrystalline structure and therefore its lack of sharp, well defined peaks.

One of the tasks of the control program is to obtain the peaks that are relevant to sample identification. In general, only two or three peaks are needed for each component. Any larger number of peaks results in an unnecessary long consultation time of LITHOS.
During a scan, data are sent from the diffractometer to the personal computer in floating point numbers. In order to reduce the influence of noise signals, an 11-point Savitzky-Golay filter is applied to the intensities and the resulting diffractogram is presented on the screen. This filter is a weighed smoothing procedure calculating each peak height from 11 terms. The filtered intensities and the scan parameters are saved on the hard disk.
If during a scan, a transmitted value surpasses a default maximum intensity, the diffractometer is instructed to restart the scanning procedure with half of the current detector sensitivity. Conversely, if the maximum intensity is too low, the scan is repeated with twice the current sensitivity.
Following the noise filtering, a baseline subtraction is performed. The baseline is a curved ascending line over the whole scanned range and results largely from background reflection from the silicon application disk, and a small 'aspecific' contribution from the sample. Numerical analysis of the baseline in a variety of diffractograms by means of the program CFIT (J.C. Hudson, 4198 Warber Dr., Flint, Michigan 48504, U.S.A.) yielded five 'best fit' equations, *e.g.*: ln[intensity] = c_1 + $c_2/[2\Theta]$. In this equation, c_1 and c_2 are

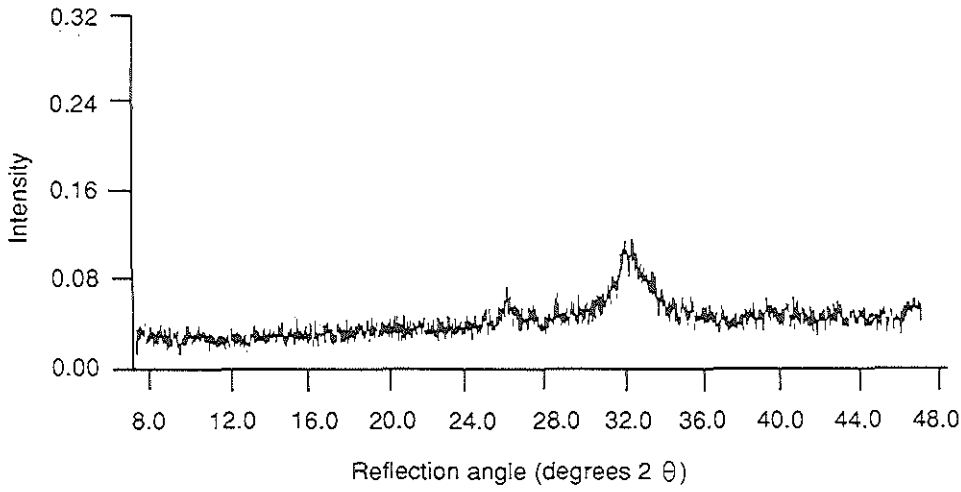


Figure 2.1 *Diffractogram of Apatite ($Ca_{10}(PO_4)_6(OH)_2$) from a renal calculus.*

constants that are estimated through least-square regression. Using this equation, the correlation coefficient of the fit to a diffractogram is typically between 0.78 and 0.86, compared with 0.69-0.76 for the fit with a straight line. The program determines the correlation coefficient for each equation and uses the one with the highest correlation for baseline subtraction.
After the baseline subtraction, the first derivative spectrum is calculated to determine the exact peak top positions. To reduce the effect of the noise which is caused by taking the derivative, the first derivative is calculated from intensities lying two steps (0.02°) instead of one step apart. From the first 250 values of this first derivative spectrum the standard deviation of the noise in this signal is calculated. In this range of the diffractogram, no peaks are encountered.
A peak entry is signalled if, in the first derivative, an intensity greater than twice the standard deviation of this noise is encountered, followed by five sequential values in ascending order. The position of the peak top is selected, if there are two adjacent non-negative values of the first derivative, followed by one negative value.

2.5 The expert system shell

For the implementation of LITHOS, the Personal Consultant™ expert system shell, version 1.10 (january 1985) from Texas Instruments Inc., Austin, Texas, U.S.A. was used. Additional use was made of IQ LISP™ version 1.7 (november 1984), from Integral Quality Inc., U.S.A.
In expert systems, the driving force of the reasoning process can be either data or conclusions. In the former case, the system tries to deduce new facts from the data it has been given (data driven reasoning or forward chaining). In the latter case the system tries to verify conclusions (goal driven reasoning or backward chaining). A combination of these types of reasoning is also possible and generally leads to more powerful reasoning processes. LITHOS however, is based on backward chaining reasoning.
The empty shell used for LITHOS is built in LISP (List Programming Language) and may be manipulated by user-made additional LISP programs. Because the Personal Consultant shell was the first commercially available version, it lacked several features which were later included in the Personal Consultant Plus shell, which is a more advanced version of its predecessor. We developed these features in the form of additional LISP programs.

2.6 Source of the knowledge

The knowledge incorporated in the expert system LITHOS has been partly obtained from an expert (diffraction patterns), partly from experiments (determination of sensitivity factors), and partly from the literature (calculation of mass percentages). The expert had

16 years of experience in diffraction analysis with the Debije-Scherrer-Hull camera and had been trained at the Faculty of Physics at the Technical University Delft, The Netherlands. X-ray diffraction analysis has been shown to perform well in both Dutch and German surveys (2).

2.7 The structure of the expert system

The knowledge base of the expert system consists of parameters, domain variables, dynamic property lists, functions, and rules. The entire system occupies approximately 315 kB of disk space. The 90 parameters are used to store information, to direct the inference process, and to drive the consultation. The 19 domain variables are used to store constants, such as 2Θ values, margins of between-run variability of 2Θ values, combinations of components that have overlapping peaks, and sensitivity coefficients relative to magnesium oxide. Property lists are generated during the consultation and serve to store temporary information that has to be manipulated, such as lists of 2Θ values and intensities, and a list of primarily matched components. The 26 LISP functions are used for a variety of purposes such as calculations, alterations in property lists, matching of 2Θ values, alterations of meta-rules (see below), production of the screen layout, storage of the result on disk.
The 102 rules are used to determine the value of parameters, to invoke LISP functions, and also to guide the inference process. The latter type of rules, called meta-rules, decide which rules are to be used in the present consultation and are constructed by corresponding LISP functions during each consultation. The rules of LITHOS are organized in 15 groups, each containing rules that functionally belong together. These rulegroups are called contexts. There is one parent context dealing with general tasks such as input, output and the selection of a list of primarily selected stone components. All other rulegroups are subcontexts of the parent context, and are designed to handle problems arising from the presence of two or more components having the same or overlapping reflection angles. One subcontext is designed to signal the presence of possible artifacts or the presence of components that are unknown to the system.

2.8 Consultation

The sequence of events in a consultation will be described in this section. For the interested reader, an extensive example is available from the author of this thesis. As mentioned above, the control program has been designed for communication with the diffractometer and for preprocessing of the raw scan data. The sequence of events is fully automated once the scan is started and can be summarized as follows:

1. Scanning
2. Storage of raw data
3. Noise filtering
4. Baseline subtraction
5. First derivative calculation
6. Calculation of the standard deviation of the first derivative
7. Peak selection
8. Storage of reduced data

The raw data from the scan are stored in order to create the possibility of repeating a consultation to study the influence of slight modifications in the control program.
The next step is the consultation with LITHOS, which is initiated by the human operator. The whole sequence is summarized below:

1. Retrieval of reduced data
2. Test for amorphous material
3. Primary selection of components
4. Sorting of components in order of decreasing dominance
5. Administration of recognized peak positions
6. Conversion of matched peak positions to default values
7. Detection and removal of artifacts
8. Correction of intensities for interference of other components and removal if necessary
9. Single component check
10. Calculation of quantitative composition
11. Administration of unmatched peaks
12. Unknown component check
13. Output of an analysis report

When the peak positions and corresponding heights have been retrieved from disk, LITHOS first checks whether more than 100 peaks have been found, in which case the sample is designated as amorphous material and the consultation is discontinued.
If this check is passed, a primary selection of components is made. In its database LITHOS contains two and sometimes three characteristic peak positions for each of 18 possible components. LITHOS compares these positions to the measured values for component recognition. For several reasons, the measured peak positions may differ slightly from the stored values and may thus fail to be recognized. To overcome this difficulty, every stored characteristic value has its specific margin within which the measured value is allowed to vary in order to still be recognized.
The next step is the sorting of the selected components in order of dominance, which is

defined as the product of the intensity of the first matched peak and the sensitivity factor for the component. This factor is a measure of the inverse reflection intensity and has been determined relative to magnesium oxide by the internal standard method of Rebentisch *et al.* (6). Later in the consultation, LITHOS will verify which components are truly present, beginning with the least dominant ones. The components that are not present are removed from the list and the remaining components are evaluated for coinciding peaks. The early removal of improbable components greatly reduces the number of coincidences that has to be checked.

After the sorting process, the measured diffraction angles of the components that have been found are compared with corresponding default values of those components. This list is analogous to the one previously mentioned but contains more peak positions and margins. The default values that have been matched and the corresponding intensities are stored in a separate list. Note that the default values and not the measured values are stored. This facilitates later retrieval procedures because the default values are invariant from run to run. The two stage process described above (selection - extensive matching) is time saving because only the selected components are compared rather than the whole list of primarily matched components.

The sorted list of components is now used to select the subcontexts of LITHOS that deal with the selected components, starting with the least dominant one. Usually the presence of these components is mimicked by minor peaks of other components which happen to coincide with the two peaks needed for recognition. As an example of a step in the clean-up procedure, the component sodium chloride will be removed if:

1. Apatite, Whitlockite, or ammonium urate is present, and
2. The intensity of the largest peak (sodium chloride) is less than 15000, and
3. The position of this component in the sorted list is third or greater.

As has been mentioned before, the removal of such 'false positive' components reduces the number of rules dealing with coincident peaks that has to be consulted.

The selection of the part of the knowledge that has to be consulted is done by a LISP function which alters the premise of one of the rules in such a manner that the inference process is guided to those rules that are relevant to the actual consultation. In this way, the content of this so-called meta-rule will be altered in each consultation. The rules dealing with components that have not been found in the sample will not be consulted and can be regarded as temporarily 'forgotten'. The option of meta-rules has been incorporated by the manufacturer in the empty shell of a newer version: the Personal Consultant Plus. The meta-rule steers the consultation along those rules that investigate the possible coincidences among components. Its premise is created from the sorted list of components in such a way that the least probable components are dealt with first. Successively, each part of this premise invokes the system to enter a subcontext dealing with one component from the sorted list. In this subcontext, a component can be removed if necessary, in a way analogous to the example of sodium chloride. For each component, LITHOS keeps a

list of all components that are known to cause coincidences with the component under consideration. This list is used to select the interfering components that are actually present. With this list a secondary meta-rule is constructed which again guides the system along all the coincidences with the component. Suppose the component ammonium urate has been detected because of coincidence of a Whewellite peak with its major peak. Then this peak can be corrected for the presence of Whewellite if a second Whewellite peak as well as the intensity ratio of both peaks in pure Whewellite are known. After the correction procedure, the component is removed from the list if the remaining intensity is below a certain limit. In this case it will be useless to correct other components for coincidence with the component that has just been removed. To avoid these fruitless efforts all other preplanned actions concerning the removed component are cancelled at the time of its removal. It is possible however, that the corrected intensity is above the removal limit. If this is the case, the component is not removed and the corrected intensity is kept by the system for further use. In this way, an intensity may be corrected more than once because of multiple interferences.
The next step is a check for a single component. The calculation of mass percentages is bypassed if, after all corrections, the sorted list contains only one component; the mass percentage of this component is then set to 100%. If there are more components, the relative composition in mass percentages is calculated according to Rebentisch (6).
At the end of the consultation, a list of measured peak positions that have not been recognized by the system is constructed. If there are many unrecognized positions, the presence of an unknown component may be suspected. This list is obtained by subtraction of all peaks associated with the diagnosed components from all the measured peaks in this sample. The former are recruted from a large list containing all peak positions that have ever been measured for each pure component (4-36 positions for each component). Finally, a screen report is generated presenting the qualitative and quantitative composition of the sample, possible warnings for artifacts like quartz, gypsum, or the presence of an unknown component.
Typical consultation times are between 30 and 60 seconds for a multiple stone, whereas stones with only one component can be analyzed within 10 seconds. In a typical example of a stone containing two components and a primary selection of five components, LITHOS used only 25% of its rules.

2.9 Conclusion

The project of developing LITHOS has taken approximately 2 years and yielded an expert system that is used in routine practice (7). Nevertheless, it should be remarked that LITHOS is not meant to function entirely without human expert supervision. In case of doubt (*e.g.* rare components, artifacts, or more than four components), the analysis is repeated with a newly powdered sample. In rare cases, an X-ray film is prepared with the

Debije-Scherrer camera in order to provide a decisive opinion. The Debije-Scherrer technique is also used for sample sizes smaller than 4 mg.
In this way we have created a situation in which the bulk of the daily routine can be dealt with by relatively inexperienced workers in the absence of a human expert. The system saves time because fewer films have to be developed and dried (14 and 10 min., respectively). The sample preparation time, the scan time, and the interpretation time are essentially the same for both methods (3, 14 and 2 min., respectively).

2.10 References

1. Boer PW, Geuns H van, Hem van der GK, Blickman JR.
Population survey in a community on the occurrence of stone disease. Proc VIIIth Congr Soc Int Urol 1979;2:64-5.

2. Röhle G.
Externe Qualitätskontrolle: Ergebnisse aus Ringversuche für Harnsteinanalysen. In: Hesse A, Claßen A, Röhle G, eds. Labordiagnostik bei Urolithiasis. 3. Seminar Urolithiasis. Stuttgart: Wissenschaftliche Verlagsgesellschaft mbH, 1988:85-92.

3. Dosch W.
Röntgendiffraktometrie. In: Hesse A, Claßen A, Röhle G, eds. Labordiagnostik bei Urolithiasis. 3. Seminar Urolithiasis. Stuttgart: Wissenschaftliche Verlagsgesellschaft mbH, 1988:53-64.

4. den Boer NC, Bakker NJ, Leijnse B.
Structuuranalyse van nier- en blaasstenen door röntgendiffractie. Ned Tijdschr Geneeskd 1972;116:373-7.

5. Blijenberg BG, Wulkan RW, Zwang L, Liem TL, Leijnse B.
Computerunterstützte Auswertung der Röntgendiffraktometrie-Analysen. In: Hesse A, Claßen A, Röhle G, eds. Labordiagnostik bei Urolithiasis. 3. Seminar Urolithiasis. Stuttgart: Wissenschaftliche Verlagsgesellschaft mbH, 1988:65-70.

6. Rebentisch G, Berg W.
Brauchbarkeitskriterien der standardisierten röntgendiffraktometrischen Harnsteinanalyse. J Clin Chem Clin Biochem 1983;21:665-71.

7. Wulkan RW, Zwang L, Liem TL, Blijenberg BG, Leijnse B.
Renal stone analysis: LITHOS, an expert system for evaluation of X-ray diffractograms of urinary calculi. J Clin Chem Clin Biochem 1987;25:719-22.

Chapter 3

LITHOS AND CALCULI

TWO COMPUTER PROGRAMS FOR THE EVALUATION OF X-RAY DIFFRACTOGRAMS OF URINARY CALCULI

3.1 Introduction

The prevalence of urinary calculi in the Dutch population is approximately 4.7% (1). With the introduction of an extracorporal shock wave lithotryptor in our hospital, the clinical chemistry laboratory has been confronted with a massive increase in requests for renal stone analysis. Traditionally, this analysis was done by X-ray diffraction with a Debije-Scherrer-Hull camera and the darkening patterns were interpreted by a specialist with many years of experience. In order to expand our capacity, a diffractometer was purchased. This instrument is equipped with an X-ray detector and it is able to communicate its output (diffraction angles and reflection intensities) to a personal computer. At the same time, an expert system shell for use on personal computers became commercially available, and we undertook the development of an expert system for interpretation of X-ray diffraction data (2). At the Faculty of Physics of the Technical University Delft and the TNO Institute of Applied Physics, Delft, The Netherlands, a Pascal program was being developed for the interpretation of digitized X-ray patterns from a Guinier-Johansson camera, of crystalline samples. A co-operation was started and the program was modified to be used in the hospital setting.
The two programs represent two different approaches to the same problem: one by means of a rule-based expert system, and one by means of a Pascal program which makes use of multivariate calibration. This prompted us to make a limited comparison of the performance of both systems.

3.2 Materials and methods

All mineralogical components used for calibration, method comparison, and reproducibility experiments were of human source. This was done because large differences in diffraction properties are known to exist between components from commercial and from human source (3). For the details of the sample preparation, diffractometer system, and diffractometer control program, the reader is referred to chapter 2, section 2.2 to 2.4.

3.2.1 LITHOS

LITHOS is an expert system for the interpretation of X-ray diffractograms of urinary calculi. The program has been described elsewhere (2). It is a rule-based, backward chaining expert system developed to calculate the qualitative and semiquantitative composition from X-ray diffractograms of urinary calculi. Further details can be found in chapter 2.

3.2.2 CALCULI

CALCULI 4.0 is a computer program for the evaluation of X-ray diffractograms of urinary calculi. It was developed at the department of Crystallography (Faculty of Technical Physics) of the Technical University Delft, The Netherlands. The program is written in Turbo Pascal 3.0 and is run on an Olivetti M24 Personal Computer equipped with a 8087 coprocessor unit. CALCULI occupies 108 kB of disk space. The program calculates the qualitative and semiquantitative composition from X-ray diffractograms of urinary calculi. CALCULI uses mass absorption coefficients from the literature for the calculation of the semiquantitative composition.
In order to calculate the qualitative composition, the program compares the peak positions of the pure components within a margin of 0.1° with the measured peak positions (the diffractogram is registrated in steps of 0.01°). The peak positions (8-29 for each component) and heights have been experimentally determined for each pure component and are contained in the program. Components are taken into account if more than two peaks have been recognized in the diffractogram.
In the calculation of the mass fractions of each component, peak heights are used to estimate the mass fraction of each component. An important difference with the approach of LITHOS is that, instead of only one peak, 8 to 29 peaks for each component are used to calculate the corresponding mass fraction. This amounts to a total of approximately 300 peak positions. The measured height at each of these 300 peak positions in the diffractogram is modelled as a linear combination of heights of the pure components that are present. The crucial point is to determine the coefficients by which the peak heights of the pure components have to be multiplied in the linear equation. This is achieved by the least squares method, where the sum of all squared differences between measured and predicted heights is minimized by varying the coefficients. When the quadratic error has been minimized, the corresponding set of coefficients is used together with the mass absorption coefficients to calculate the relative mass concentrations of all components.

3.3 Results

3.3.1 Method comparison

In order to investigate the merits of the respective computer programs, a comparison was made with the results of the Debije-Scherrer method (Table 3.1). For a number of renal calculi, the relative content of one component as found by LITHOS or CALCULI was compared with the content found by the Debije-Scherrer method.

Table 3.1 *Regression of LITHOS (method 1) or CALCULI (method 2) on the Debije-Scherrer method. The number of observations N, correlation coefficient R, slope, intercept, and the standard deviation of the residual error of regression S_{yx} are presented.*

Component	Method	N	R	Slope	Intercept (%)	S_{yx} (%)
Whewellite	1	150	0.98	0.97[a]	3[b]	5
	2	40	0.97	0.96	4	6
Weddelite	1	93	0.98	0.97[a]	2[b]	4
	2	31	0.91	0.73[a]	2	8
Apatite	1	142	0.95	0.98	1	6
	2	33	0.84	0.86	8[b]	10
Struvite	1	56	0.97	1.00	−1	5
	2	5	0.99	1.15	−5	3
Brushite	1	35	0.97	1.00	0	5
Uric acid	1	26	0.98	1.04	−4	4
Uric acid dihydrate	1	10	0.99	1.05	0	3
Ammonium hydrogen urate	1	5	0.87	0.94	5	10
Whitlockite	1	9	0.93	1.09	−9	8
Cystine	1	5	1.00	0.93	7	1

[a]Significantly different from 1 (Student t test; $p < 0.05$)
[b]Significantly different from 0 (Student t test; $p < 0.05$).

3.3.2 Reproducibility

In order to investigate the reproducibility of the analytical method in combination with both computer programs we measured material from one renal stone fifteen times. The means and standard deviations are shown in Table 3.2. Standard deviations were not significantly different between LITHOS and CALCULI (Fisher's exact test; $p < 0.05$).

In several further experiments we tried to quantify the contribution to the total variation of the effects of different steps in the sample preparation.
Experiment a. Five times a mixture of 40% Whewellite, 40% Weddelite, and 20% Apatite was prepared, powdered, and a diffractogram was taken. This experiment was designed

Table 3.2 Reproducibility example of LITHOS (method 1) and CALCULI (method 2). Fifteen measurements of the same human calculus. For explanation see text.

Component	Method	Mean(%)	Stand. dev.(%)
Whewellite	1	44	6.4
	2	39	4.5
Weddelite	1	42	4.9
	2	36	3.0
Apatite	1	14	6.0
	2	24	5.9

to measure the combined effects of powdering and application. The variability introduced by weighing is negligible.

Experiment b. A mixture was made of the above composition, powdered. With this mixture, five diffractograms were produced, *i.e.* powder of exactly the same composition was applied five times to the silicon disk and analyzed. This was done to measure the effect of sample application.

Experiment c. A mixture of the same composition was applied once to the silicon disk and five spectra were taken with removal and replacement of the disk between the runs. This was done to estimate the influence of possible disturbance of the powder layer.

Experiment d. Five spectra of the same mixture were prepared without opening of the diffractometer between the runs. This was done to exclude variation that is due to sample handling. This experiment was repeated twice to confirm the rather remarkable results. The results of these experiments with both systems are shown in Table 3.3.

Next, two experiments were done to exclude effects induced by the diffractometer as a source of variation. In the first experiment we powdered a sample of quartz, applied and measured it in six successive runs without intermittent disk removal. The intensity of the

Table 3.3 *Reproducibility experiments (40% Whewellite, 40% Weddelite, and 20% Apatite). a. Sample prepared, powdered and applied five times; b. Sample prepared and powdered once and applied five times; c. Five runs of a sample with intermittent disk removal; d1-d3. Five runs of a sample without intermittent disk removal. Method 1: LITHOS, Method 2: CALCULI.*

Component	Method	Mean (%)						Stand.dev. (%)					
		a	b	c	d1	d2	d3	a	b	c	d1	d2	d3
Whewellite	1	38	45	37	38	42	37	2	2	1	2	2	3
	2	38	38	38	45	39	24	6	4	3	4	8	2
Weddelite	1	34	38	53	50	39	38	3	1	2	1	2	4
	2	34	32	36	39	36	40	6	3	2	4	8	3
Apatite	1	28	17	10	12	18	25	3	1	2	1	3	4
	2	28	31	26	16	25	36	10	6	4	8	15	4

major peak was measured and the variation coefficient was calculated to be 8.4%. In a control experiment we measured in the same manner a quartz sample that had been fixed to the silicon disk and was obtained from the manufacturer. The variation coefficient of the major peak was 0.9%.

3.3.3 Unknown component

In the evaluation phase of CALCULI it became apparent that the combination of Whewellite and the presence of an unknown component frequently occurred. Examination of the Whewellite base diffractogram showed a shift of 0.02-0.04° for all diffraction angles when compared with a newly measured diffractogram of pure Whewellite. The latter was installed as the new Whewellite base. To investigate the effect of this alteration, the diffractograms of 20 renal stones were selected which contained both Whewellite and the presumed unknown compound. Ten of these were binary, five were ternary and another five were quaternary mixtures. The percentage of unknown component in these stones varied between 8 and 19%. After correction of the Whewellite base, 18 stones were found to contain regular components only, whereas 2 stones remained that still contained the unknown compound. In both, the percentage of unknown component had shown a decrease to 7%. The quantitative compositions of 17 out of 20 stones had come closer to those found by LITHOS.

3.3.4 Learning about ammonium hydrogen urate

One of the components that are rarely found in renal stones in Western Europe is ammonium hydrogen urate. A visiting scientist from Indonesia, confronted us with the problem of calculating ammonium hydrogen urate. He brought a large number of renal stones that almost invariantly contained ammonium hydrogen urate. This enabled us to teach the expert system to recognize the new component. Table 3.4 shows the results of LITHOS and the Debije-Scherrer method.

Table 3.4 *Composition (%) of a calculus containing ammonium hydrogen urate.*

Debije-Scherrer		LITHOS[1]		LITHOS[2]	
Amm. hydr. urate	50	Uric acid. $2H_2O$	21	Amm. hydr. urate	75
Weddelite	40	Weddelite	41	Weddelite	25
Whewellite	10	Whewellite	trace	Whewellite	trace
		Apatite	28	Whitlockite	trace
		Whitlockite	10		

[1]Original consultation
[2]After modification.

3.3.5 Consultation time

Although it is not an essential item, we have made a comparison of the consultation times with both systems (Table 3.5). Typical consultation times of 50 seconds are obtained with Whewellite/Weddelite/Apatite, which is a frequently occurring composition. LITHOS can do pure components within 10 seconds, but this is only a slight advantage compared to the sample preparation time (3 minutes) and the scan time (14 minutes). Occasional outliers in consultation time up to two minutes occur in both systems. Not included in Table 3.5 is the time needed for baseline subtraction and peak selection, which is approximately half a minute for both systems.

Table 3.5 *Consultation time (seconds).*

Qualitative composition	LITHOS	CALCULI
Cystine	7	30
Whewellite	9	33
Apatite	11	30
Whewellite/Weddelite	48	47
Struvite/Apatite	54	45
Brushite/Whitlockite	54	120
Uric acid/Apatite	94	70
Whewellite/Weddelite/Apatite	52	55
Brushite/Weddelite/Apatite	47	55
Whitlockite/Weddelite/Apatite	54	65

3.4 Discussion

3.4.1 Method comparison

The results obtained with LITHOS and CALCULI have been compared to those obtained with the Debije-Scherrer method, in which the X-ray films are judged by a human expert (Table 3.1). Although this seems a rather subjective approach, it was the best alternative because X-ray diffraction analysis is regarded as being the golden standard and our in-house experience had proven its virtues in previous external quality control programs.
From Table 3.1 it can be concluded that both systems perform reasonably well. CALCULI has a relatively small slope for Apatite because it has the tendency to overdiagnose Apatite when present in smaller amounts. This effect tilts the regression line and thus gives a low slope. A secondary effect is caused by the frequent co-occurrence of Apatite and Weddelite. The overdiagnosing of Apatite results in underdiagnosing of Weddelite when it is present in large amounts. By analogy this gives a small slope. The reason for this lies in the microcrystalline structure of Apatite, which produces a very broad peak

of low intensity. When Apatite is present in small amounts it becomes difficult to distinguish the peak from the baseline noise.

3.4.2 Precision

In order to evaluate the precision of both methods, a stone of frequently occurring composition was taken and analyzed fifteen times. Table 3.2 shows standard deviations within 7%. Standard deviations may of course vary with absolute composition. However, from the regression figures associated with Table 3.1 we can conclude that for the three most frequently occurring compounds: Whewellite, Weddelite, and Apatite, only low amounts of Apatite will show large imprecision. Therefore, we chose to use the composition mentioned above, and to try and trace the major source of imprecision in this setting.
A well known effect in X-ray diffraction analysis causing imprecision and inaccuracy is preferential orientation, which means that crystal planes are not exposed in a random fashion to the radiation. This causes some peaks in the diffractogram to become higher and others to become smaller than normal, which in turn leads to over- or underestimation of the mass concentration. In the Debije-Scherrer method the random orientation is provided by continuous rotation of the sample in the X-ray beam. However, in the method with a fixed silicon applicator this is not the case, so great care has to be taken to prevent preferential orientation.
To dissociate the effects of the different steps in the sample preparation we devised the experiments shown in Table 3.3. Unfortunately, no large differences were observed between the steps of powdering, application, and disk removal. Perhaps the most striking effect is the imprecision resulting from successive runs without opening the diffractometer. When this experiment was repeated, similar results were obtained. For instance, the peak intensity of Whewellite at 15.03° changed from 24478 to 19759 within five runs without opening the diffractometer, which is as much as 19% reduction. In a similar experiment with a fixed quartz sample we obtained a variation coefficient of 0.9% for the intensity of the highest peak, compared to 8.4% with a powdered quartz sample. This led us to conclude that the contribution of the diffractometer to the total variation is negligible.
We can only speculate about the cause of the disturbing effect mentioned above, which hampers every attempt to improve the technique of the sample preparation. One possibility is the building up of static electricity in the instrument. This is known to concentrate at sharp points of conductors and might influence the orientation of the microparticles in the X-ray beam.

3.4.3 Accuracy

The evaluation of the accuracy in renal stone analysis by X-ray diffraction is a difficult

task. The technique of X-ray diffraction is regarded as the most reliable in this field. The accuracy of this method can not be evaluated by means of artificial samples prepared from commercially available components because these are known to have crystal structures that are different from their natural counterparts. Therefore, the accuracy of such methods can only be obtained by secondary means, *e.g.* round robin tests with mixtures of natural components of maximal purity.
The comparison of two programs judging the same diffraction data is an even more difficult task. Some idea can be obtained from the comparison with the Debije-Scherrer method (Table 3.1). The slopes for most of the components are satisfactorily, although CALCULI displays some overestimation of Apatite, especially in the lower percentages (figure 3.1).
This compound presents a difficulty because the diffractogram of pure Apatite is highly variable in shape (figure 3.2) and therefore only a grand mean can be used.
The accuracy of LITHOS is diminished by the variation resulting from preferential orientation. Because LITHOS relies for its calculation on only two peaks for each component, the phenomenon of preferential orientation is liable to exert considerable influence on the results.

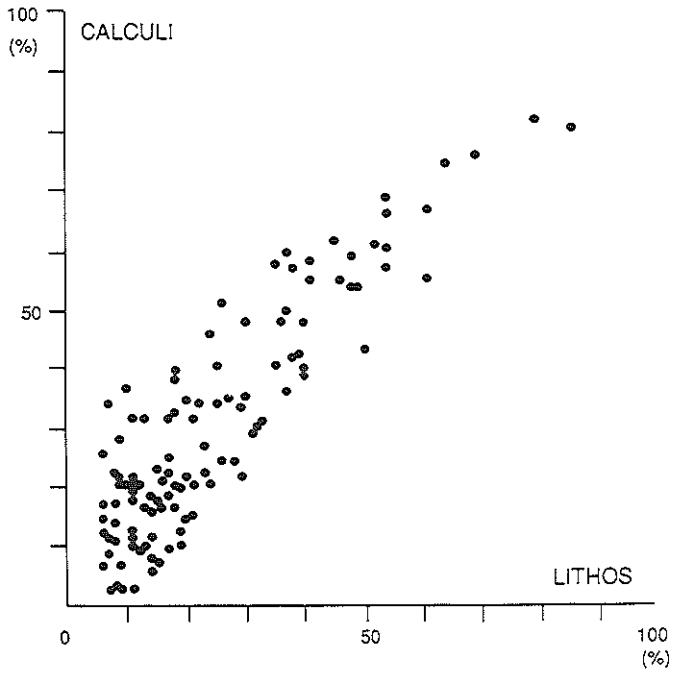
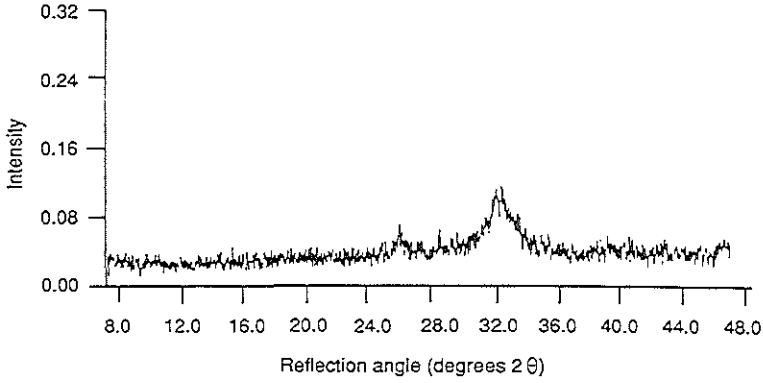
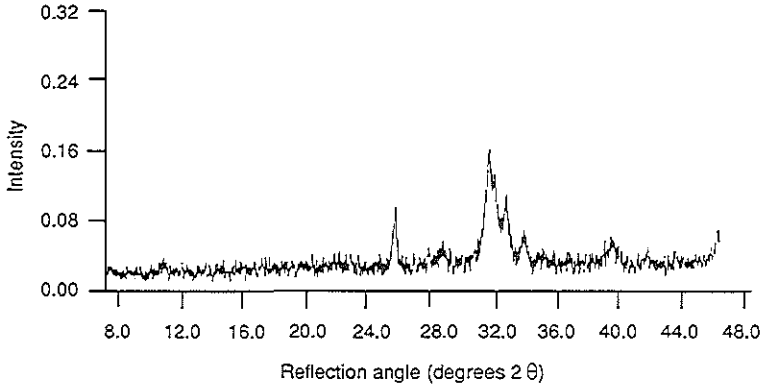


Figure 3.1 *Comparison of the results of LITHOS and CALCULI for Apatite.*

1.



2.



3.

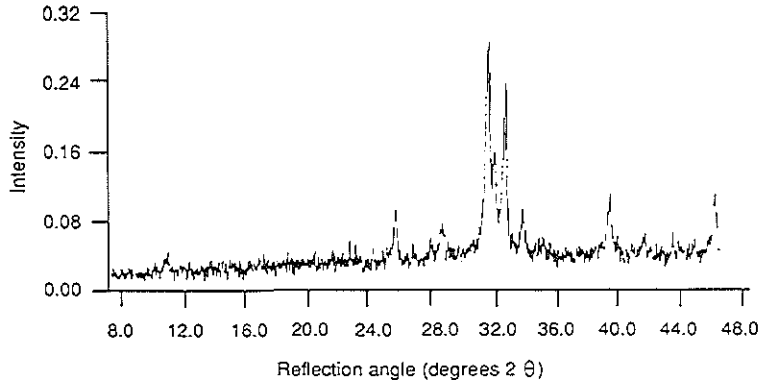


Figure 3.2 *Diffractograms of Apatite from: 1. A renal calculus (human), 2. Commercial source, and 3. A molar (human).*

3.4.4 Additional comments

The comparison of two computer programs for interpretation of the same set of data is of course a precarious undertaking, especially because there is no hard-core gold standard in this field. The results found by us represent a momentum in the ongoing development towards perfection. Both programs have their advantages and drawbacks. The use of an expert system shell has the advantage of enabling the inexperienced program developer to make a prototype in relatively little time, compared with the use of a conventional programming language. As has been shown above, the quality of the results these systems produce is in this case determined to a large extent by factors concerning the sample preparation.
Therefore, in the daily routine, both programs are used side by side to judge each sample. If the results show a consensus, then they are taken to be correct and are transferred to the physician. If there is a disagreement, then the human expert examines the input data for preferential orientation. Often, a consensus is found when such a sample is redistributed on the disk and rerun. Results from external quality surveys over a period of two years (12 samples) have been qualitatively correct in all cases, quantitatively correct in 10 cases and deviating by 10% quantitatively in only two cases.

3.5 References

1. Boer PW, van Geuns H, van der Hem GK, Blickman JR
Population survey in a community on the occurrence of stone disease. Proc VIIIth Congr Soc Int Urol 1979;2:64-5.

2. Wulkan RW, Zwang L, Liem TL, Blijenberg BG, Leijnse B
LITHOS, an expert system for evaluation of X-ray diffractograms of urinary calculi. J Clin Chem Clin Biochem 1987;25:719-22.

3. Blijenberg BG, Wulkan RW, Zwang L, Liem TL, Leijnse B
Computerunterstützte Auswertung der Röntgendiffraktometrie-Analysen. In: Hesse A, Claßen A, Röhle G, eds. Labordiagnostik bei Urolithiasis. 3. Seminar Urolithiasis. Stuttgart: Wissenschaftliche Verlagsgesellschaft mbH, 1988:53-64.

Chapter 4

INTERPRETATION OF ACID-BASE MEASUREMENTS

4.1 Terminology and interpretation of acid-base disorders

The interpretation of acid-base parameters should begin by a clear definition of the terminology. The debate about this subject, now 25 years ago, has created two schools. These two schools hold different interpretations of the same terminology, which are liable to cause much confusion if definitions and interpretations have not been clearly defined.

The term 'acidosis' may serve as an example to explain this disparity. The first school, often said to use the 'laboratory interpretation', applies this term to describe the chemical condition of blood at one point in time. This condition is defined as a pH below the lower reference limit. It should be stressed that this is a static description and therefore cannot be used to decide by which process or processes this condition is reached.
The other school, using the 'clinical' or 'physiological interpretation', defines acidosis as a process, leading to an accumulation of acid (or loss of base) in the body. This is a dynamic interpretation and therefore cannot be used to describe the actual measured value at one point in time.
Nevertheless, the term 'acidosis' is used very often to describe the clinical condition of a patient based on a measurement at one point in time. Here, a new element is introduced, *i.e.* the fact that not only the acidity of the blood, but also other analytes are measured in the same sample. These are the carbon dioxide tension (PCO_2) and the bicarbonate concentration [HCO_3^-] in the blood. The measured value for both entities is the result of the rate of production and the rate of excretion from the body. All three variables are subject to an interdependency which is described by the Henderson-Hasselbalch equation.

$$pH = pK' + \log\frac{[HCO_3^-]}{S \times PCO_2}$$

In this equation, pK′ is the negative logarithm (base 10) of the apparent equilibrium constant of the reaction between carbon dioxide and water to give a bicarbonate ion and a hydronium ion. The constant S relates the concentration of dissolved carbon dioxide to its partial pressure in equilibrium with the solution.

In order to make a correct acid-base diagnosis, knowledge is needed of the physiological regulation mechanisms such as the body's response to disturbances in the acid-base equilibrium, *i.e.* the increase in bicarbonate production by the kidney, which is a slow process compared with the removal of carbon dioxide by the lungs, and the ventilatory response, which is dependent on the interstitial pH in the brainstem.

The knowledge about physiological mechanisms has led to the use of the adjectives 'metabolic' and 'respiratory'. Again, two different interpretations can be given of these terms.
According to the laboratory interpretation, a metabolic acidosis is regarded as a decreased concentration of the metabolic component in the blood. In contrast, the physiological interpretation of metabolic acidosis is a condition resulting from an accumulation of acid (other than CO_2) or from loss of bicarbonate. Once again, the laboratory interpretation describes the state of the blood, whereas the physiological interpretation describes the process.

Two other qualifications that are often used in acid-base diagnosis are 'simple' and 'mixed'. These are used to denote clinical conditions that result from one or from several etiological factors, respectively. Thus, a patient who has uncontrolled diabetes may have a metabolic acidosis with a ventilatory response and yet have a simple acid-base disorder. Would there be coexisting pulmonary insufficiency, then the condition would be denoted as a mixed acid-base disorder. It will be evident that these qualifications are not used in the laboratory interpretation.

All disturbances of the acid-base equilibrium give rise to a compensatory response of the body in an attempt to maintain homeostasis. The degree of this compensation can be measured in two ways. In the laboratory interpretation, the compensation is said to be complete if the blood pH has been restored to normal. For the simple acid-base disturbances, the amount of compensation in bicarbonate or carbon dioxide occurring normally is known and can be calculated with the corresponding confidence limits (1,2). The physiological interpretation denotes a compensation as being complete if the amount of response is within these confidence limits, regardless of the blood pH.
It may be difficult, even in simple disturbances, to recognize which is the primary process and which is the compensatory process. When a patient with a metabolic acidosis is artificially ventilated, a ventilatory overcompensation may be the result, leading to a slightly elevated pH. Without knowledge of the clinical situation, one could be misled to conclude that a primary respiratory alkalosis was present, accompanied by a secondary and partial metabolic compensation. The same could occur with a patient taking medication which stimulates the respiratory center. Two examples will be given to further illustrate the complexity of the acid-base interpretation.

Case 1

A 70-year-old man who had been persistently vomiting over several days was admitted to the hospital with congestive heart failure (3). On examination he was hyperventilating. The arterial pH was 7.58, the PCO_2 was 2.8 kPa (21 mm Hg), and the bicarbonate concentration was 19 mmol/L. The corresponding reference ranges are 7.35–7.45, 4.65–5.98 kPa (35–45 mm Hg) and 22–32 mmol/L, respectively. This condition could be described as a simple respiratory alkalosis (with partial metabolic compensation). At this point, a new element in acid-base diagnosis is introduced, *i.e.* the anion gap (the sum of the sodium and potassium – minus the sum of the chloride and bicarbonate concentrations). The anion gap was elevated (33 meq/L, ref. range 7–17 meq/L) in the absence of any medication or therapy likely to increase the unmeasured anions, suggesting that metabolic acidosis was present. Because the anion gap is increased by 16 meq/L above the upper reference limit, with normal sodium, potassium, and chloride concentrations, a similar decrease in bicarbonate would be expected. However, the bicarbonate concentration was decreased by far less than 16 mmol/L, suggesting a 'hidden' underlying metabolic alkalosis. This finding was consistent with the history of persistent vomiting. Thus, a triple acid-base disorder was present, *i.e.* respiratory alkalosis, metabolic acidosis, and metabolic alkalosis. Without the clinical information, the only way to detect this multiple disorder is by calculation of the anion gap.

Case 2

A 60-year-old man who developed acute oliguric renal failure (4), was admitted in a confused state to the hospital after several days. The arterial pH was 7.19, the PCO_2 5.0 kPa (38 mm Hg), and the bicarbonate concentration 14 mmol/L. This could be described as a simple metabolic acidosis, which is consistent with the elevated anion gap (27 meq/L) and the clinical history. In this case, the decrease in bicarbonate is reflected in an equal rise of the anion gap. Normally, this situation would trigger a respiratory compensation causing the PCO_2 to fall. In this patient, an expected PCO_2 of approximately 3.8 kPa (29 mm Hg) can be calculated. The actual level (5.0 kPa) suggests an underlying respiratory acidosis, which is most likely due to concurrent congestive heart failure. Thus a double disorder was present, *i.e.* metabolic acidosis and respiratory acidosis. Without clinical information, the only clue to the solution is the calculation of the normal amount of compensation.

In these examples, a correct (physiological) acid-base diagnosis can be made which is based solely on the blood-gas values, the anion gap, and a knowledge of the normal amount of compensation. However, without the clinical information (artificial ventilation, medication) it is often impossible to conclude which is the primary disturbance and which

one is secondary. This in turn could lead to an erroneous calculation of the compensation and therefore to faulty conclusions. Some researchers have therefore constructed two sets of acid-base interpretations, one for the normal situation, and one for patients on a ventilator (5,6).

Another problem in acid-base diagnosis is the choice of an indicator of metabolic changes. Commonly, the actual bicarbonate concentration is used, but this approach is not altogether correct, since this concentration is not independent of the carbon dioxide tension. Indicators that are independent of the PCO_2 are the standard bicarbonate concentration and the base excess.
The standard bicarbonate concentration is the actual bicarbonate concentration which has been normalized to a standard PCO_2 of 40 mm Hg. The base excess is a measure of the total buffering capacity of the blood. It is defined as the base concentration in meq/L of whole blood as measured by titration with strong acid to pH 7.40 at a PCO_2 of 40 mm Hg at 37 °C when the hemoglobin is 100% saturated with oxygen (7). For negative values of the base excess, the titration must be carried out by strong base.
The base excess is slightly dependent on the hemoglobin concentration and it can be used to calculate the amount of bicarbonate or acid needed to correct an acid-base disturbance. We have calculated the base excess at two hemoglobin concentrations and several combinations of pH and PCO_2. The results, which are summarized in Table 4.1, show that at normal PCO_2, the base excess varies only slightly with the hemoglobin concentration. In the cases with a low pH, where bicarbonate administration is usually considered, the highest variation was found with high PCO_2 (8 kPa). A difference of 2.48 meq/L (roughly 0.35 meq/L for each mmol Fe^{2+}/L Hb) was found, while the base excess at a hemoglobin concentration of 10 mmol Fe^{2+}/L is 10.0 meq/L. This condition is likely to occur less frequently than a low pH with a low PCO_2, in which case the variation was only 0.91 meq/L with a base excess of −21.5 meq/L.
The use of the base excess has been advocated by Scandinavian researchers, whereas American researchers tend to use the bicarbonate concentration (8). This may give rise to different interpretations, as can be seen from the following example.

Table 4.1 *The absolute difference (δBE, in meq/L) between the base excess values at two fixed hemoglobin concentrations (3 and 10 mmol Fe^{2+}/L) as a function of acidity (pH) and carbon dioxide pressure (PCO_2, in kPa).*

pH	7.40	7.40	7.40	7.55	7.55	7.55	7.15	7.15	7.15
PCO_2	5.32	2.70	8.00	5.32	2.70	8.00	5.32	2.70	8.00
δBE	0.01	1.61	1.96	0.10	2.42	2.80	1.46	0.91	2.48

Case 3

A 55-year-old man with carcinoma of the lung complicated by a superior vena cava syndrome was transferred from another hospital for emergency radiotherapy (9). At the referring hospital he had required intubation for a respiratory arrest that occurred shortly after admission. Just before transport he was reported to have 'normal blood gases'. During transport, the ventilation was maintained manually using an inspired relative oxygen content of 100%. On admission, his arterial pH was 7.04, the PCO_2 13.9 kPa (104 mm Hg), the bicarbonate concentration 27 mmol/L, and the anion gap was 17 meq/L with normal electrolytes. This disorder can be described as a simple respiratory acidosis. The bicarbonate concentration is essentially the same as the expected 28 mmol/L, calculated for an acute respiratory acidosis (2). Assuming a hemoglobin concentration of 9 mmol Fe^{2+}/L, a base excess of approximately −8 meq/L (ref. range −3 to +3 meq/L) and a standard bicarbonate concentration of 18 mmol/L (ref. range 21–25 mmol/L) can be calculated. Considering these values, and the anion gap which was at its upper reference limit, a slight element of metabolic acidosis was present. The high content of inspired oxygen abolished the hypoxemic drive, and thereby reduced the alveolar ventilation. Unfortunately, no lactate determination had been done. Thus, a double disorder was present, *i.e.* respiratory acidosis and metabolic acidosis.

This case clearly demonstrates the advantages of using an indicator of metabolic changes which is less dependent on the PCO_2 than the bicarbonate concentration. It should be admitted that the hemoglobin concentration is of some influence on the base excess: a concentration of 6 mmol/L would have given a base excess of −5 meq/L. Much criticism has been given to the use of this calculated base excess, because the CO_2 titration curve of blood *in vitro* differs widely from that of blood *in vivo* which is, to a high degree, dependent on intracellular buffering (10). In our opinion, this argument carries less weight than the dependency of the bicarbonate concentration on the respiration. Therefore, we have chosen to use the base excess as an indicator of metabolic disturbances (11).

The interpretation of acid-base disorders is determined by only three parameters, *i.e.* the pH, PCO_2, and base excess and it is essentially the same as the one given by Vallbona *et al.* (12). However, some modifications have been made.

1. We have refrained from using the term 'compensation', because there is no way of knowing whether the patient has artificial ventilation or some other non-physiological stimulus to ventilation. Thus, a low pH, a low PCO_2, and a low base excess would be diagnosed as a metabolic acidosis and respiratory alkalosis instead of a metabolic acidosis with some degree of respiratory compensation.
2. The term 'mixed' is not used because it refers to the existence of multiple etiological factors.

3. A condition with low pH, normal PCO_2, and low base excess is diagnosed as a metabolic and respiratory acidosis instead of a metabolic acidosis. The normal PCO_2 in this case points to an absence of the normal respiratory reaction to a metabolic acidosis. The absence of this reaction is causing a respiratory acidosis, regardless whether its cause is airway obstruction, drug overdose, or artificial underventilation.
4. We have included terms for the deviation of only one acid-base parameter, in accordance with the recommendations of the *ad hoc* committee (13).

In theory, 27 combinations can be made out of three parameters with three possible states (Table 4.2). Of these, 1 combination represents the normal situation, 12 are regular acid-base disorders, 6 are deviations of only one parameter, and the remaining 8 do not occur and have been termed 'non-denominable'.
Although we feel that polarizing between 'laboratory' and 'clinical' interpretations creates an unsound atmosphere, our diagnosis is nevertheless based on laboratory results only. This approach has often been attacked as being unrealistic, but we believe that it is yet a valid one, provided that the physiological basis and the restrictions to its application are kept firmly in mind.
We have not yet included the use of the anion gap into this system of acid-base diagnosis. The reasons for this are multiple and depend largely on organizational factors. Analysis of blood gases and of serum electrolytes are done on different equipment, in different material, and at different times of the day. The results can be combined for evaluation either in a laboratory database or in a hospital information system. The decision to do data analysis at the location where the data are produced is a matter of

Table 4.2 *Proposed interpretation of acid-base parameters. Abbreviations are the same as in Table 4.1. The signs −, n, and + denote values below, within, and above the reference range, respectively.*

pH	PCO_2	BE	Interpretation
Any	−	−	Metabolic acidosis and respiratory alkalosis
−	n/+	−	Metabolic and respiratory acidosis
−	+	n	Respiratory acidosis
n	n	n	Normal
+	−	n	Respiratory alkalosis
+	n/−	+	Metabolic and respiratory alkalosis
Any	+	+	Metabolic alkalosis and respiratory acidosis
−	n	n	Acidemia
+	n	n	Alkalemia
n	−	n	Hypocapnia
n	+	n	Hypercapnia
n	n	−	Hypobasemia
n	n	+	Hyperbasemia
Any other condition			Non-denominable

communication efficiency and favors the use of a laboratory database. Furthermore, it should be known if the time of sampling is the same, which very often is not the case. Arterial punction is done by the medical staff, whereas venapuncture is done by the nursing staff. The arterial samples are brought on ice to the laboratory and are analyzed as soon as possible, but the venous samples are part of the large number of requests for analysis which is daily received. For this reason, the electrolyte values are known at a later time of day than the blood-gas values. It is this time of day which ultimately determines the moment when a full acid-base diagnosis can be made.
New blood-gas equipment has been developed, which also measures electrolytes in the same blood sample. Besides the advantage in timing, there is also the advantage of using the same material. For potassium, a difference between serum and plasma has been reported, which results from the coagulation process (14). Unfortunately, most of these new instruments can analyze sodium and potassium but not chloride, thus preventing the calculation of the anion gap. Favourable exceptions are the instruments from NOVA biomedical and Ciba Corning.

4.2 Evaluation of the proposed interpretation

Some researchers have criticised the use of the base excess as being often misleading in acid-base diagnosis (15). Some examples used in this criticism are shown in Table 4.3.

Table 4.3 *Acid-base interpretation. Abbreviations are: HCO_3 = actual bicarbonate concentration, BE = actual base excess, M. and m. = metabolic, R. = respiratory, Ac = acidosis, Alk = alkalosis, comp = compensation. Missed diagnoses using the interpretation of Table 4.2 have been underlined.*

	pH	PCO_2 mm Hg	HCO_3 mmol/L	BE meq/L	Interpretation Schwartz	This thesis
1.	7.32	80	40	11	R.Ac (m.comp)	M.Alk, R.Ac
2.	7.18	80	29	0	R.Ac (m.comp), M.ac	R.Ac
3.	7.12	80	25	-5	R.Ac (m.comp), M.Ac	M.Ac, R.Ac
4.	7.41	80	49	23	R.Ac (m.comp), M.Alk	M.Alk, R.Ac

1. This interpretation can be considered to be correct.
2. In this case, there are two opposite metabolic disturbances, leading to a normal base excess. Here, knowledge of the anion gap would have been useful to unmask the metabolic acidosis. It can also be detected by calculating the expected bicarbonate concentration for this chronic respiratory acidosis (37 mmol/L), which is 8 mmol/L higher than the actual level of 29 mmol/L. Comparison of the anion gap and the change in bicarbonate concentration could have reveiled the compensating metabolic alkalosis.
3. In this case, only the compensatory metabolic alkalosis is missed. To prevent this, the same solution as given for the second example can be applied.
4. This interpretation can be considered to be correct.

There are several other well-documented cases in the literature (3,9,16,17). When these cases are considered, the necessity of knowing the anion gap is apparent once more. For example, in the 22 cases reported by Cohen and Kassirer, 16 were diagnosed correctly; of the remaining 6 cases, 4 would have been diagnosed correctly if the anion gap could have been used; one case had inconsistent data, and one case was a borderline problem. The use of only one metabolic parameter does not permit the simultaneous recognition of two metabolic disturbances (acidosis and alkalosis).
By these examples, the importance of measuring the anion gap is emphasized once more. The above proposal for acid-base interpretation without the use of the anion gap is not a definitive one, but should only be regarded as a primary step towards a more complete system.

4.3 The need for systems offering acid-base interpretation

1. In a study, 71% of the physicians questioned believed they did not require computer assistance for acid-base diagnosis (18). When actually tested, however, they were able to correctly identify only 39% of the acid-base problems, including mixed disturbances.
2. In another study which was done by three emergency physicians, the correct response rates for single, double, and triple disorders were 86%, 49%, and 17%, respectively (19). The normal situation was recognized in only 68% of the cases, whereas an included laboratory error was recognized in 19% of the cases. The questionnees were physicians at various levels of training from emergency medicine, internal medicine, pediatrics, surgery and family practice specialties.
3. In a third study, a system for acid-base disorders was compared with the opinion of a 'medical expert', which was used as a gold standard (2). Disagreement in only 22 out of 194 clinical cases (11%) was found, of which 14 were borderline cases. In the remaining 8 cases, the physician had to admit that in fact the systems interpretation was right.
4. In a fourth study, a system for disorders in fluid and electrolyte balance was developed that was superior to the junior clinicians in the identification of acid-base disorders (20).

4.4 Systems for interpretation of acid-base disorders

The development of computer programs for acid-base interpretation is now at least twenty years old (5). Cohen *et al.* combined the Siggaard-Andersen nomogram with the knowledge of the physiological response to acid-base disturbances in order to determine the acid-base status of a given patient. They calculated the percentage of physiological compensation provided that it is known whether the patient is on or off a ventilator. The data were entered on punched cards into an IBM 360 computer, programmed in FORTRAN IV. Since then, many researchers have developed computer programs for acid-base interpretation, in some cases including electrolyte disturbances (Table 4.4). Instead of

presenting a lengthy discussion of all publications, we will highlight some characteristics of these interpretative systems.

4.5 Characteristics of interpretative systems

The list in Table 4.4, which may not be complete, contains mainly systems for acid-base interpretation although some systems have been designed for fluid and electrolyte management as well.
The hardware used varies widely in capacity, ranging from a simple pocket calculator to a powerful workstation. Systems that only do some calculations and comparisons may well be implemented on calculators with intermediate capacity, whereas the larger systems that apply forms of reasoning may need more powerful computers (workstations). Some of these systems are used on a time-sharing basis and can be consulted by

Table 4.4 *Systems for the interpretation of acid-base disorders. n.r. means not reported.*

Programming language	Hardware	Remarks	Lit. ref.
FORTRAN	IBM 360/75		5
CAL	n.r.		21
n.r.	IBM 7094 II		22
n.r.	IBM 360/50		12
STRcomp, MUMPS, APL	PDP-9/PDP-1D	1	23-26
BASIC, COURSEWRITER	DP-15/GE-635		23-26
n.r.	Prog.101/Olivetti		27
FORTRAN	PDP-6		28
n.r.	Prog.101/602/Olivetti		29
BASIC	Radio Shack TRS-80	2	30
LISP	LISP machine	2, 3	31, 32
n.r.	n.r.		33
BASIC	HP 41C pocket		34
n.r.	IBM-XT, AT/Apple II/IIe	1, 4	35
BASIC	MPT/100		36
BASIC	n.r.		6
PASCAL	Apple		37
BASIC	Apple II		38
n.r.	IBM-PC	5	39
BASIC	IBM-PC		19
PROLOG	SUN-3/160		2, 40
n.r.	Intel280/310	2, 6	20

1. Available from the author
2. Additional interpretation of electrolytes
3. A(cid) B(ase) EL(ectrolytes)
4. Blood Gas Program
5. Model Fluid Acid Base
6. Shell Guru.

telephone (23). The programming languages that are used show a wide variability as well. Many of the programs have been written in BASIC, some in FORTRAN, there is one program in PROLOG (2), and one program applies a relational database (39). The input required by such systems may be entered interactively (keyboard) or automatically (on-line). All systems listed in Table 4.4 require interactive input.
An important characteristic of an interpretative system is whether it uses clinical data or not. It has long been held that no interpretation is possible without clinical data, but nevertheless 6 programs from Table 4.4 can do without it (2,5,22,27,30,37).
The interpretative systems can be classified according to their general approach. The simpler systems are designed to do some straightforward calculations and comparisons. Most of them were developed at a time that many calculations had to be done by hand or by the use of nomograms, whereas in the modern analyzers many of these calculations have been incorporated. In addition, there is one system that is model-based (39). The complex systems use some sort of reasoning. There is one system which uses PROLOG, one of the artificial intelligence languages.
Another characteristic is the purpose for which these systems have been developed. Broadly speaking this may be either training or consultation assistance. Some purposes that have been mentioned are: backup function for nights and weekends when expert opinion is not readily available, early warning, assistance in small hospitals, assistance in busy laboratories, intensive care and emergency departments, and education on the job.
Another characteristic is the amount of appreciation that is encountered with the introduction of a system. Although the published data may have a favourable bias, the overall picture is positive. Some citations are: 'patient care favourably influenced', 'well appreciated by the medical staff', 'at times outperforms its inventor'. An overview of the psychological factors in introducing these systems is given in the literature (15). Data about the evaluation phase of these systems is only scarcely available. In three publications, greater than 90% agreement is claimed with medical expert opinions. From an evaluation of 255 cases a specificity of 94% and sensitivity of 99% have been found (2).

4.6 Development of an expert system for acid-base and electrolyte disorders

In view of the well-defined problem area and the need for interpretative support, it was decided to develop an expert system CHEMPATH for acid-base and electrolyte disorders. It was developed using the shell Personal Consultant Plus from Texas Instruments. This is a newer version of the shell used for LITHOS (cf.ch.2). When the consultation is started the user is presented with a screen layout where all acid-base and electrolyte parameters can be entered. This approach is more efficient than the time-consuming question-and-answer approach where the data are asked for and entered sequentially. A first version of CHEMPATH was developed in co-operation with Dr. B.K. van Kreel at the Department

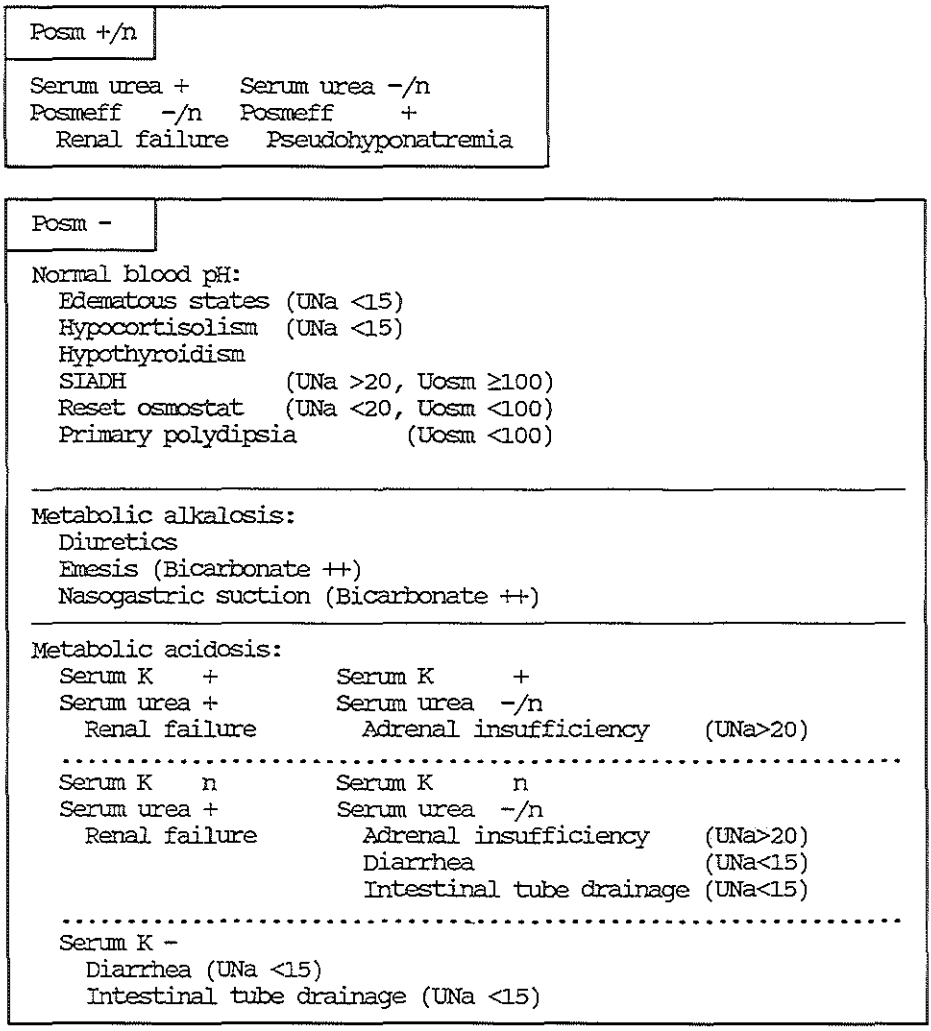
of Chemical Pathology, Erasmus University Rotterdam, Rotterdam, The Netherlands. CHEMPATH is designed to deal with three types of input: blood-gas data, electrolyte data, or a combination of these. There are rule groups for acid-base analysis, for hypo- or hypernatremia, and for hyponatremia or hypokalemia in combination with blood-gas measurements. Apart from the serum sodium or potassium concentration, other analytes may be required such as serum chloride (to calculate the anion gap), or the urine osmolality. This creates another difficulty in the application of such a system. Blood-gas data and electrolyte values may have been measured, but the urine osmolality may be unknown (this is known as the problem of missing data). In such cases, CHEMPATH gives a list of all diagnoses consistent with the input values, regardless of the value of the missing analyte. In this sense it is a restrictive system: the more data are known, the shorter the list of conclusions that are still consistent. The decision schemes in CHEMPATH have been largely constructed from knowledge extracted from ref. 1, with some slight modifications. Figure 4.1 gives an example for the decision scheme for hyponatremia in combination with blood-gas data. There are, as has been mentioned, more decision schemes but the one shown in figure 4.1 is a representative example of the general approach.

Conditions of hyponatremia should first be classified by serum osmolality. If this is decreased, a subdivision can be made by acid-base condition. Hyponatremia with a decreased osmolality and an accompanying metabolic acidosis can be further classified by serum potassium, serum urea, and urine sodium concentrations. The findings between parentheses are those that are usually found with the corresponding disorder.
This approach is of course subject to criticism. It will be of no use to a physician to have to request serum sodium, osmolality, and blood-gas values before arriving at the diagnosis 'emesis'. Many of the other diagnoses will be apparent from their clinical symptoms. Another difficulty is the fact that the system is limited to single diagnoses, and is unfit to advise in mixed electrolyte disorders. This is a classical problem, often referred to as 'the problem of the exploding search space'. CHEMPATH however, may be useful to those who want to test their knowledge in the course of their training in *e.g.* clinical chemistry or internal medicine. Additionally, it may serve to demonstrate how the spectrum of possibilities can be narrowed by a carefully aimed request sequence. As has been said earlier, there is also the organizational problem of collecting the data, and finally it must be admitted that the LISP shell is relatively slow in the consultation process.

These factors have led us to limit ourselves to the acid-base diagnostic part, and to translate this part into Turbo Pascal. In daily routine, blood-gas analyses are done on ABL 330 equipment from Radiometer, Copenhagen, Denmark. The data are sent to a personal computer which is coupled via a local area network to the hospital information system. Because the software for the data communication has been written in Turbo

Pascal, the acid-base program has been translated into the same language.
Before this new course was taken, experiments had been done with a laboratory database in DBASE III. The data from analytical instruments which are coupled to a local area network were accumulated in a central database for interpretation by an expert system.

Figure 4.1 *Decision scheme for hyponatremia. An explanation is given in the text. Abbreviations are: Na = Sodium (mmol/L), K = Potassium (mmol/L), U = urinary, Posm = serum osmolality (mosmol/kg), Posmeff = effective serum osmolality, SIADH = Syndrome of Inappropriate AntiDiuretic Hormone secretion, +, n, and - = above, within, and below the reference range.*



The expert system shell Personal Consultant Plus is equipped with functions which enable it to read a DBASE III file. This file, in turn, was filled by a Turbo Pascal program included in the application software of the analytical instrument. However, this solution proved to be very slow and impractical compared to data handling that was completely regulated by a Turbo Pascal program.
When a blood-gas analysis has been done, the analyst may generate a printed report with the sample number, the blood-gas data, and the laboratory diagnosis. This report may be handed over to the nurse who has brought the sample and has been instructed to ask for such an interpretation. Alternatively, a code representing the diagnosis may be sent to the hospital information system and be translated for those departments that are interested in this service. Because the venipunctures are done by the nursing staff, the use of bar-code labels has not yet been introduced in our hospital. This means that only the sample number and not the patient number is available to the personal computer and the network, nor is the time of sampling. This makes the problem of combining data from different apparatus a very difficult one. It is hoped that in the future the organizational aspects of the laboratory will be changed in such a way as to enable a practical and direct data interpretation.

4.7 Conclusions

In conclusion it can be said that the incorporation of a system for diagnostic support in acid-base and electrolyte disorders in a hospital laboratory is dependent on several factors. The most important of these is the simultaneous availability of data from different analytical equipment. This can be achieved by retrieval from a laboratory database, or by retrieval from the hospital information system. A system for the interpretation of acid-base disorders has been proposed, and it has been shown that the anion gap would be a valuable adjunct. An expert system for diagnostic support in acid-base and electrolyte disorders has been developed, which, although not fully tested, has produced encouraging results with cases used for examination of the trainees in clinical chemistry.

4.8 References

1. Rose BD
Clinical physiology of acid-base and electrolyte disorders. 2nd ed. New York: McGraw-Hill Book Company, 1984.

2. Pincé H, Verberckmoes R, Willems JL
Computer-aided interpretation of acid-base disorders. Int J Biomed Comp 1990;25:177-92.

3. Walmsley RN, White GH
Mixed acid-base disorders. Clin Chem 1985;31:321-5.

4. Puschett JB, Greenberg A, eds.
Disorders of fluid and electrolyte balance. Diagnosis and management. London: Churchill Livingstone Inc., 1985.

5. Cohen ML
A computer program for blood-gas analysis. Comp Biomed Res 1969;2:549-57.

6. Kirkeby OJ, Risöe C
Computer interpretation of acid-base status. A practical way of dealing with a difficult problem. Br J Clin Pract 1985;39:377-8.

7. Siggaard-Andersen O, Engel K
A new acid-base nomogram. An inproved method for the calculation of the relevant acid-base balance of human blood. Scand J Clin Lab Invest 1960;12:177-86.

8. Winters RW
Definitions and terminology in blood acid-base chemistry. Ann NY Acad Sci 1966;133:211-24.

9. Cohen JJ, Kassirer JP
Acid-base. 1st ed. Boston: Little, Brown and Company Inc., 1982.

10. Howorth PJN
The physiological assessment of acid-base balance. Br J Dis Chest 1975;69:75-97.

11. Siggaard-Andersen O, Wimberly PD, Fogh-Andersen N, Goethgen IH
Measured and derived quantities with modern pH and blood gas equipment: calculation algorithms with 54 equations. Scand J Clin Lab Invest 1988;48 (suppl 189):7-15.

12. Vallbona C, Pevny E, McMath F
Computer analysis of blood gases and of acid-base status. Comp Biomed Res 1971;4:623-33.

13. *Ad hoc* comittee on acid-base terminology
Report of *ad hoc* committee on acid-base terminology. Ann NY Acad Sci 1966;133:251-8.

14. Wulkan RW, Michiels JJ
Pseudohyperkalaemia in thrombocythaemia. J Clin Chem Clin Biochem 1990;28:489-91.

15. Schwartz WB, Relman AS
A critique of the parameters used in the evaluation of acid-base disorders. New Engl J Med 1963;268: 1382-8.

16. Walmsley RN, White GH
Normal "anion gap" (hyperchloremic) acidosis. Clin Chem 1985;31:309-13.

17. Narins RG, Emmett M
Simple and mixed acid-base disorders: a practical approach. Medicine 1980;59:161-87.

18. Hingston DM, Irwin RS, Pratter MR *et al.*
A computerized interpretation of arterial pH and blood gas data: Do physicians need it? Resp Care 1982; 27:809-15.

19. Schreck DM, Zacharias D, Grunau CFV
Diagnosis of complex acid-base disorders: physician performance versus the microcomputer. Ann Emerg Med 1986;15:164-70.

20. Autio K, Kari A, Tikka H
Integration of knowledge-based system and database for identification of disturbances in fluid and electrolyte balance. Comp Meth Prog Biomed 1991;34:201-9.

21. Schwartz WB
Medicine and the computer. The promise and problems of change. New Engl J Med 1970;283:1257-64.

22. Suero JT
Computer interpretation of acid-base data. Clin Biochem 1970;3:151-6.

23. Bleich HL
Computer evaluation of acid-base disorders. J Clin Invest 1969;48:1689-96.

24. Bleich HL
The computer as a consultant. New Engl J Med 1971;284:141-7.

25. Bleich HL
Computer-based consultation. Electrolyte and acid-base disorders. Am J Med 1972;53:285-91.

26. Bleich HL
Computerized clinical diagnosis. Fed Proc 1974;33:2317-9.

27. Rowberg A, Lee S
Use of a desk-top calculator to interpret acid-base data. Am J Clin Pathol 1973;59:180-4.

28. Goldberg M, Green SB, Moss NL, Marbach CB, Garfinkel D
Computer-based instruction and diagnosis of acid-base disorders. A systematic approach. J Am Med Assoc 1973;223:269-75.

29. Grogono AW
Assessment of acid-base disturbances employing a desk-top computer. Br Med J 1973;1:381-4.

30. Swezey CB, Jacobson W
Computer-based diagnostic reporting for serum electrolytes. Am J Clin Pathol 1980;74:812-9.

31. Patil RS, Szolovits P, Schwartz WB
Causal understanding of patient illness in medical diagnosis. Proc 7th Int Joint Conf Artif Intell 1981;2:893-9.

32. Patil RS, Szolovits P, Schwartz WB
Modelling knowledge of the patient in acid-base and electrolyte disorders. In: Szolovits P, ed. Artificial intelligence in medicine. Boulder, Colorado: Westview Press, 1982:191-226.

33. Sapir DG, Hoeck DL, Dandy WE
Computer diagnosis of simple and mixed acid-base disorders. Crit Care Med 1981;9:266.

34. Kievit J
Standardized diagnosis and treatment of fluid, acid-base and electrolyte disorders in the surgical patient with the aid of a programmable pocket calculator. Br J Surg 1983;70:282-5.

35. Goldstein RA
Review of "Blood Gas Program". Ann Intern Med 1985;103:483.

36. Kaldor G, Rada R
Computerized evaluation of acid-base disoders based on a nine-cell decision matrix. Med Biol Eng Comp 1985;23:269-73.

37. Maude DL
Acid-base disorders - a computer simulation. Physiologist 1985;28:118-21.

38. Skaredoff MN
A computerized system for rapid interpretation of acid-base disorders. Int J Biomed Comp 1986;18:229-38.

39. Flood RL, Cramp DG, Leaning MS, Carson ER
Mathematical modelling of fluid-electrolyte, acid-base balance for clinical application. In: Salamon R *et al.*, eds. MEDINFO 86. Elsevier Science Publishers B.V. (North-Holland), 1986:133-7.

40. van Uytven H, Verberckmoes R, Willems JL
Een expert systeem in PROLOG voor zuur-base evenwichtsstoornissen. Proc 7th Med Inf Congr. de Moor G, Stevens C, eds. Antwerpen, Belgium. 1987:121-6.

Chapter 5

MULTIVARIATE REFERENCE REGIONS

5.1 Introduction

The theory of multivariate analysis for the determination of reference ranges has as yet found limited application. For an introduction to multivariate analysis the reader is referred to the work of Albert and Harris (1). Thus far, this theory has been used for the calculation of multivariate reference ranges from healthy populations (see further), for multivariate delta checks (2,3), and for discriminant analysis.

As has been stated in chapter 1, the use of multivariate reference ranges is most useful with variables that are appreciably correlated. In this chapter, only Gaussian distributions will be discussed. For an n-dimensional Gaussian distribution, the volume V containing 95% of all observations is given by (4):

$$V = \frac{(\chi^2(n)_{0.05})^{n/2}\ \pi^{n/2}}{\Gamma\ (n/2 + 1)} \prod_{1}^{n} (\sigma_i)\ \sqrt{P} \qquad \text{(eq. 1)}$$

where χ^2 denotes the Chi-square function, σ_i is the standard deviation of the i-th distribution, Γ is the gamma function, and P denotes the determinant of the correlation matrix $|r_{ij}|$, where r_{ij} is the correlation coefficient between the i-th and the j-th variable. The volume V is represented by an ellipsoid in hyperspace. For the bivariate case this is an ellipse, and equation 1 reduces to $5.99\ \pi\ \sigma_1\ \sigma_2\ \sqrt{(1 - r^2_{1,2})}$. This means that for a pair of variables with correlation coefficient 1, the area is zero (straight line). The area is maximal for a correlation coefficient 0 (circle). Therefore, the difference in area between a double univariate reference region and a bivariate reference region increases with increasing correlation coefficient; this also holds for higher dimensionalities. It is this difference in area which is one of the important properties of multivariate analysis.

For an univariate Gaussian distribution, a fraction 0.95 of all observations is found within two standard deviations (2σ) of the mean value, encompassing a range of 4σ. If two independent univariate distributions are considered, the combined reference region can be visualized as a rectangle. If the sides of the rectangle have a length of $4\sigma_1$ and $4\sigma_2$ (the indexes representing the variates), then a fraction of $0.95^2 = 0.90$ of the observations will be located in this rectangle. If a fraction 0.95 ($= 0.975^2$) is desired, then the sides must be enlarged to $4.48\sigma_i$. In this case the area within the rectangle is equal to $4.48^2\ \sigma_1\sigma_2$.

According to equation 1, the area within the bivariate reference ellipse is 5.99 π $\sigma_1\sigma_2$ for uncorrelated variables. This is 94% of the area of the combined univariate rectangle. For three uncorrelated variables this percentage decreases to 84%. For any given dimensionality, this fraction decreases with an increasing correlation coefficient.

In a two-dimensional distribution four types of observations can be distinguished (Figure 5.1):

1. normal both univariate and bivariate
   (inside rectangle and ellipse)
2. abnormal both univariate and bivariate
   (outside rectangle and ellipse)
3. bivariate normal yet univariate abnormal
   (inside ellipse, outside rectangle)
4. univariate normal yet bivariate abnormal
   (inside rectangle, outside ellipse)

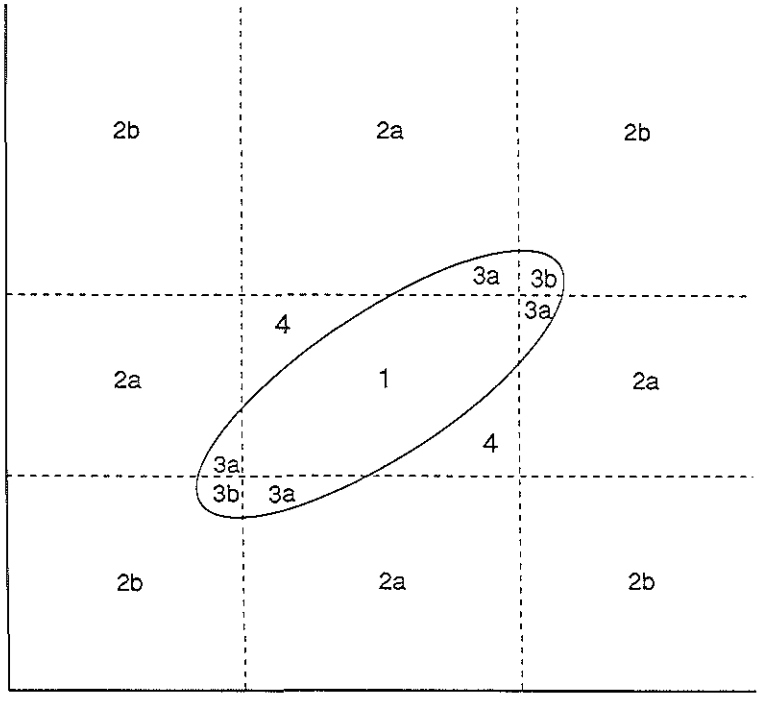


Figure 5.1 *Classes of observations with two measurements.*

The larger the correlation coefficient, the larger the area available for category 4 measurements. It depends entirely on the behaviour of both variables in pathology if it is likely that values in this area are found. If this is the case, a gain of diagnostic space has been attained (clinical advantage). If not, a gain of space for plausibility control has been attained (laboratory advantage).
Application of the general formula to the trivariate case yields for the volume V of the 95% ellipsoid (2):

$$V = 91.425\ \sigma_1\sigma_2\sigma_3\ \sqrt{(1 - r^2_{12} - r^2_{13} - r^2_{23} + 2r_{12}r_{13}r_{23})} \qquad \text{(eq. 2)}$$

5.2 Multivariate reference regions obtained from healthy populations

Several authors have reported correlations between clinical chemistry variates obtained from healthy individuals. Table 5.1 gives some examples of the highest correlation coefficients.

The reduction in reference area can attain high values. For the combination of total lactate dehydrogenase (LD), heat stable LD, and aspartate aminotransferase, a 50% smaller reference range was reported, as compared to the univariate region, even with correlations not exceeding 0.42 (10).

Table 5.1 *Absolute correlation coefficients (r) between, and the number of observations (n) of clinical chemistry variates, obtained from healthy individuals.*

Variables	r	n	Lit. ref.
LD - HBD	0.97	125	5
Chloride - Total CO_2	0.75	250	5
ALAT - ASAT	0.72	1979	6
$PavCO_2$ - HCO_3^-	0.72	25	7
Total protein - Albumin	0.65	313	8
Total calcium - Albumin	0.62	1979	6
CSF IgG - CSF Albumin	0.60	127	9
Sodium - Uric acid	0.58	1979	6
Direct bili - Indirect bili	0.54	205	10
Uric acid - Creatinine	0.54	284	11

Abbreviations are: LD = lactate dehydrogenase, HBD = hydroxybutyric acid dehydrogenase, CO_2 = carbon dioxide, ALAT = alanine aminotransferase, ASAT = aspartate aminotransferase, $PavCO_2$ = arteriolized venous carbon dioxide tension, HCO_3^- = plasma bicarbonate, CSF = cerebrospinal fluid, IgG = immunoglobulin G, bili = total bilirubin.

5.3 Advantages and disadvantages of the multivariate approach

The question as to under which conditions multivariate reference regions are most fruitfully applied, remains to be answered. For this, an inspection of the advantages and disadvantages of the multivariate approach is necessary.
Multivariate reference ranges have originally been developed to reduce the high proportion of falsely abnormal results as seen with profiles of multiple (univariate) laboratory tests. As has been stated in chapter 1, the number of falsely abnormal results rapidly increases with the number of tests under consideration. This proportion is usually assumed to be $1 - (0.95)^n$ where n is the number of variates (6,12), but this only holds for independent variates, which is usually not the case.
The relative insensitivity of the multivariate model to univariate abnormality of a single test has been evaluated (6). It has been shown that this insensitivity increases with the number of tests under consideration. In contrast, univariate deviations of tests which are highly correlated with other tests (*e.g.* ASAT with ALAT) are reflected much earlier in the multivariate result than deviations of tests that are not highly correlated with others (*e.g.* glucose). In other words, increasing the number of tests has a dampening effect on the sensitivity to single deviations, and the model is more sensitive to single deviations with tightly correlated tests.

5.4 Some examples

Some authors have reported the fraction of observations that are univariate abnormal and multivariate normal (category 3, section 5.1) relative to the total number of results. Fractions of 4% have been reported for albumin and total protein in a population of patients (13) and 9% for a three-parameter thyroid profile in a population of healthy individuals (14). The probability that such observations occur has been shown to be 1.4% for tests with pairwise correlations of 0.6 (15). The same author has shown that this probability is relatively insensitive to the number of variables, up to five.
In the opposite case (category 4), where the results of a number of tests are univariate normal and multivariate abnormal, the area where such observations may be located increases with the correlation coefficient. Applying equation 1 to the data from ref.13, an area of $5.99\ \pi\ \sigma_1\sigma_2\ \sqrt{(1 - 0.52^2)}$ can be calculated for the ellipse, and an area of $4.48^2\ \sigma_1'\sigma_2'$ for the rectangle. If the standard deviations were equal, the area of the ellipse would be 80% of the area of the rectangle. Naus *et al.* found that 1.1% of the albumin and total protein measurements, obtained from a population of patients were in this category (correlation coefficient 0.52, ref. 13). Apparently, hardly any pathology that could lead to category 4 type of observations was present in this population. This in turn creates the possibility to use the multivariate approach in plausibility control. Category 4 results may be found because of pathology (low probability in this case) or by

Table 5.2 *Correlation coefficients as obtained from 'healthy' populations, in cases where multivariate analysis has been used to construct a reference region.*

Tests	r			n	Lit. ref.
$PavCO_2$ - HCO_3^-	0.72			25	7
Total calcium - Albumin	0.59			313	8
CSF IgG - CSF Albumin	0.60			127	9
LD - LDh - ASAT	0.42	0.38	0.16	200	10
FT_4I - FT_3I - ln TSH	0.24	−0.19	−0.09	3752	14

Abbreviations are the same as in Table 5.1. LDh = heat-stable LD, FT_4I = free thyroxine index, FT_3I = free triiodothyronine index, ln TSH = natural logarithm of thyrotropin.

laboratory error in one or more of the measurements. Hence the possibility of detecting laboratory errors may be increased by rerunning all category 4 results.

In conclusion, multivariate reference regions are best used to reduce the number (and follow-up measurements) of false-positive univariate results in 'healthy' persons (laboratory advantage) and to increase the sensitivity for early detection of pathology (clinical advantage).
In Table 5.2, some examples are given of cases where multivariate reference regions have been constructed from 'healthy' populations.

5.5 Assumptions underlying the calculation of reference ranges

In the preceding sections, the multivariate reference ranges as obtained from healthy individuals have been discussed. The use of test results obtained from diseased individuals to develop univariate reference ranges was already known, and has been subject to heavy criticism (16,17). This criticism has been aimed at the use of the Gaussian model, the stability of the data, and the validity of the assumptions underlying the use of patient data:

1. The distribution around the mode is Gaussian.
For many tests this has been shown not to be the case, even in healthy populations. Hence, this assumption should be verified in the calculation procedure. Several methods for such tests exist; the result of this verification depends on the sensitivity of the statistical test that is used (18). If the distribution is not Gaussian, a transformation of the data is usually indicated to satisfy the above assumption.
2. The Gaussian part of the distribution is stable and unaffected by variations in disease composition, age, etc.
This of course cannot be guaranteed, and therefore these ranges will have to be regularly adjusted. Population shifts have been reported in the univariate approach (19,20).

3. The analytical accuracy and precision have been stable in the period of data collection. The validity of this assumption can be checked by means of quality control data of the laboratory where the data were produced.
4. The resulting range reflects the range of 'healthy' results.
In general, this is not true. It has been shown repeatedly that patient-based reference ranges are shifted away from the 'healthy' mean, and generally display larger standard deviations (16,21-23). Of course it should be remembered that the early developed methods for estimating such ranges were coarse (24) and have gradually been refined (21,25).
Only this last point of criticism is specific for the use of patient data. The other points apply equally to the use of data from healthy populations, which is common practice in contemporary clinical chemistry. It will be clear that the above criticism also applies to the determination of multivariate reference regions from patient populations.

Concerning the first point of criticism, according to Pèpe *et al.* the use of a Gaussian distribution as a model is justified if (26):

1. The measured value is influenced by many factors.
2. The fluctuations caused by these factors are independent.
3. The fluctuations have the same order of magnitude.
4. The frequency of large fluctuations is low.

For clinical chemistry measurements in general, the conditions 1, 2 and 4 seem to hold (18). The third condition, however, is not always valid. The activity of an enzyme in serum, for example, is largely governed by its rate of 'production' and its clearance from the circulation. In pathological processes, the spillover rate is usually much higher than the clearance and therefore condition three is no longer valid. The reverse is true for albumin: the rate of loss in renal failure may largely exceed the rate of synthesis by the liver.
It seems therefore reasonable to assume that variables which are regulated both at high and low values, are more liable to follow a Gaussian distribution than variables that are easily either 'gained' or 'lost'. This comment applies to the mode and the standard deviation of the patient distribution. These values are usually shifted away from those obtained from healthy persons.
In general, patient-derived standard deviations are larger than their equivalents obtained from healthy populations. However, ideally, intra-individual reference ranges from healthy persons should be obtained. These represent a refinement in that they are smaller than the ranges obtained from a healthy reference group as a whole, and hence offer an earlier detection of disease. Madias *et al.* have done a first step in this direction in a multivariate analysis of the plasma bicarbonate concentration and the arteriolized venous carbon

dioxide pressure (7). They used a series of intra-individual measurements to construct a bivariate reference region, thereby eliminating the inter-individual variation from the reference distribution.

5.6 Multivariate reference regions from patient populations

Naus *et al.* were the first to calculate a bivariate range from a collection of patient data (13). They used a large number of test results (3840 pairs) and determined the fraction that followed a Gaussian distribution. The characteristics of the reference ellipsoid were estimated by the method of Bhattacharya (27,28). They found a mean value of 71.9 g/L for total protein, 44.0 g/L for albumin, and standard deviations of 4.87 and 3.13 g/L, respectively, and a correlation coefficient of 0.49. Unfortunately, they did not report their conventional univariate reference limits (mean ± 2 s.d.), but their results would lead to univariate 95% reference limits of 62–81 g/L and 38–50 g/L. This is in reasonable agreement with the 60–78 g/L and 35–50 g/L cited in a leading textbook (29).
It must be emphasized that the multivariate patient-based reference region does *not* serve to detect if a patients results are normal (*i.e.* comparable to 'healthy' individuals), but it is devised to:
1. Achieve a reduction in unnecessary follow-up of marginally deviating results (cost saving).
2. Detect either early pathology or measurement error (clinical advantage and quality improvement of laboratory results). This is the case when observations are univariate normal, yet multivariate abnormal.
3. Serve as a measuring device to compare the patients results to the whole group of patients, *i.e.* to identify those patients who seem to require most attention.

5.7 Preliminary investigation of patient data

The study of acid-base terminology described in chapter 4 prompted us to investigate the distribution of acid-base diagnoses over a hospital population during a period of one month (30). In 1350 cases, the normal condition (pH, PCO_2, and base excess within univariate reference limits) was found most frequently (32% of the cases, see figure 5.2). The most dominant acid-base disturbance was an uncompensated respiratory alkalosis (20% of the cases). In 21% of the cases, only one parameter was outside its univariate reference range. All other disturbances occurred with frequencies less than 8%.
When the group with only one abnormal parameter was analyzed in more detail, a decreased PCO_2 was found most frequently (48%), followed by an elevated base excess (27%). All other single deviations occurred in percentages lower than 9%.
A striking feature is the large proportion (21%) of single deviations, which could at the time not be classified in terms of proper acid-base diagnoses. A similar phenomenon with

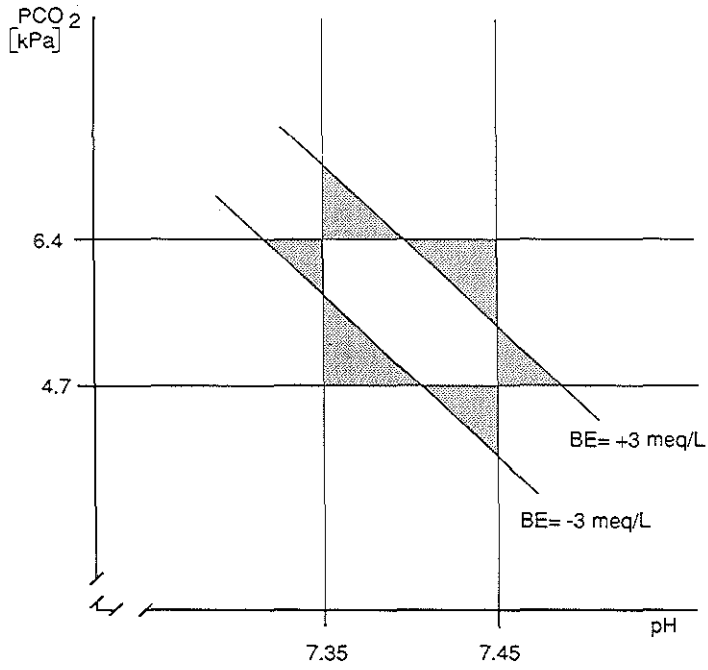


Figure 5.2 *Siggaard-Andersen diagram with grey areas of near-normal results.*

acid-base measurements has been reported elsewhere (31): In a series of 255 cases, 20% of the results were between 2 and 3 standard deviations of the univariate reference ranges. These data were denoted as equivocal, *i.e.* they could neither indicate nor exclude the presence of an acid-base disorder. The authors commented that follow-up results were no longer equivocal, *i.e.* the condition had either improved or deteriorated. Another remarkable similarity with our results was the high frequency of respiratory alkalosis (20%, the amount of compensation was not reported). We observed a frequency of 20% for uncompensated respiratory alkalosis.
Adressing the same problem, other authors have used two concentric ellipses, for bicarbonate and PCO_2: one ellipse defining the normal region as stated by Madias *et al.*, the other arbitrarily defining the 'near-normal' cases (32). Details describing the ellipses were not reported.

5.8 Theory of the multivariate reference region

An n-dimensional Gaussian distribution is completely defined by the means and the standard deviations of the n variables and the n(n−1)/2 correlation coefficients. In an

n-dimensional space, the *loci* of the points of equal probability density are ellipsoids given by:

$$(x - \mu)^T C^{-1} (x - \mu) = \text{constant}$$

In this equation, μ is the true mean vector and C is the variance-covariance matrix. In general, μ and C are unknown and must be estimated from a set of observations. If **m** and S are the estimates of μ and C, one may define:

$$d_i^2 = (x_i - \mathbf{m})^T S^{-1} (x_i - \mathbf{m})$$

as the non-Euclidian distance of an observation vector x_i to the estimated mean **m** (33). As the equation indicates, points with equal distance to the mean are located on an ellipsoid.
The probability of occurrence of d^2 will be denoted by $P(d^2)$. This probability follows an F-distribution with N and n degrees of freedom, where N is the number of observations from which the parameters of the distribution were estimated and n is the dimensionality. For large N, the F-distribution may be approximated by a Chi-square distribution with n degrees of freedom.
For the bivariate case the ellipsoid reduces to an ellipse. If σ_1 and σ_2 denote the standard deviations for the respective variables, and $\sigma_1 > \sigma_2$, then the angle Θ between the major axis of the ellipse and the x-axis is given by:

$$\tan \Theta = \frac{r\ \sigma_1\ \sigma_2}{\sigma_1^2 - \sigma_2^2}$$

Thus, the orientation of the ellipse is determined by the correlation coefficient r and the standard deviations σ_i. If each variable is normalized by subtraction of the mean and division by the standard deviation, the ratio of the lengths of the major and minor axis is given by the expression: $\sqrt{(1 + |r|)} / \sqrt{(1 - |r|)}$ which is equal to unity for a correlation coefficient zero. Graphically this is represented by a circle; non-scaled variables with unequal standard deviations would still produce an ellipse.

5.9 Estimation of the multivariate reference region

A procedure to estimate the characteristics of the distribution from a sample of observations in a n-dimensional case will now briefly be summarized. The details of this procedure can be found elsewhere (34). Starting point is a first approximation of the vector of the means and of the covariance matrix, *e.g.* as obtained from the total sample. The sample is then trimmed to those observations with a d^2 smaller than an arbitrarily chosen limit. In the publication by Gelsema *et al.*, a value of 4 is used (34). From these,

new values for the mean and for the elements of the covariance matrix are then estimated in an iterative process. The standard deviations are corrected for the effect of trimming and it is verified whether the distributions of d^2 of the subsample is compatible with a Chi-square distribution. If it is, the subsample is enlarged to the observations with $d^2 <$ 4.5 and the estimation procedure is repeated. In subsequent steps the threshold value for d^2 is increased by 0.5. The whole procedure is discontinued when the subsample is no longer compatible with a Gaussian distribution. The last obtained compatible estimates then define the Gaussian kernel of the total population.
The number of observations needed to produce reliable results is considerable. It depends on the dimensionality of the model and the number, location and symmetry of the aberrant observations in the distribution.

5.10 Multivariate reference regions for acid-base results

With the procedure as summarized above, we have calculated reference ellipsoids for pH, PCO_2 and base excess from unselected data from the population in two hospitals (Table 5.3). The data from the childrens hospital were mainly obtained from arteriolized capillary samples, whereas the data from the adult group were measured in arterial samples. No corrections for patient temperature were made.

Table 5.3 *Characteristics of the reference ellipsoids for pH, PCO_2, and base excess (BE) for several age groups. Abbreviations are: n = number of observations, s.d. = standard deviation, r = correlation coefficient, d = days, m = months, y = years.*

	n	pH mean s.d.	PCO_2 mean s.d. (kPa)	BE mean s.d. (meq/L)	r pH, PCO_2	r pH, BE	r PCO_2, BE
Newborn (<5d)	2261	7.307 0.049	5.813 0.863	−3.732 2.164	−0.715	0.549	0.187
Infant1 (5d–1m)	2646	7.301 0.045	6.395 0.849	−2.358 2.666	−0.541	0.601	0.344
Infant2 (1m–1y)	2810	7.357 0.041	5.653 0.950	−1.010 2.944	−0.514	0.349	0.621
Child1 (1–5y)	1365	7.381 0.041	4.853 0.568	−2.173 2.537	−0.409	0.687	0.378
Child2 (5–18y)	1072	7.387 0.038	5.013 0.525	−1.210 0.246	−0.512	0.672	0.285
Adult (>18y)	2402	7.438 0.044	4.770 0.780	0.240 3.120	−0.490	0.580	0.410

In the adult population, 75% of the patients were classified as multivariate normal in contrast to 26% classified as univariate (conventionally) normal for the three variates. This means a threefold reduction in apparent pathological values in the multivariate

approach. The 'reference volume' in the univariate case is the volume of a rectangular box with sides $2 \times 1.96\ \sigma_i$, calculated to be $60.2\ \sigma_1\sigma_2\sigma_3$. From equation 2 (section 5.1), the volume of the multivariate ellipsoid can be calculated to be $13.7\ \sigma_1'\sigma_2'\sigma_3'$. If the standard deviations were equal, this would represent a more than threefold smaller reference space. Usually, patient-derived standard deviations are larger than their health-based counterparts, and this will reduce the difference in reference volumes.
When the distribution of acid-base disorders over this data set was investigated, uncompensated respiratory alkalosis, concluded with the univariate procedure, was found most frequently (in 20% of the cases, consistent with the percentage found in an earlier set of 1350 cases, section 5.7).
When only the observations outside the ellipsoid were used (*i.e.* multivariate abnormal results), this percentage decreased to 2%. In this multivariate abnormal group, uncompensated respiratory acidosis was found with the highest frequency (5%), followed by uncompensated metabolic acidosis (4%).
In 20% of the total number of cases there was only one parameter outside its univariate reference limit. This category was called non-denominable in a univariate approach (section 5.7). By definition, in the multivariate approach, all observations are denominable.

5.11 Discussion

With respect to the results, several remarks can be made. For the adult population, there is an obvious shift in reference ranges (Table 5.4).
The population means are shifted towards higher pH and lower PCO_2, consistent with the high proportion of respiratory alkaloses in the hospital population. The standard deviations are approximately twice as large as those conventionally used (health-based). Larger ranges for patient-based populations have been reported before (*e.g.* 13,22). In another, smaller sample from the hospital population similar results were obtained (35). Currently, the characteristics of the reference ellipsoid in the population from another hospital is being examined.

Table 5.4 *Means and standard deviations of reference regions for adults.*

	Multivariate (patient-based)		Univariate (health-based)	
	mean	s.d.	mean	s.d.
pH	7.44	0.044	7.40	0.025
PCO_2 (kPa)	4.77	0.780	5.55	0.433
BE (meq/L)	0.24	3.120	0	1.530

The frequent occurrence of respiratory alkalosis can be explained in several ways. The majority of the patients is artificially ventilated. According to clinical experience, these patients tend to feel better with PCO_2 values slightly lower than those in healthy individuals. They are thus kept in a state of slight overventilation. Furthermore, a situation of mild uncompensated respiratory alkalosis has been reported in anemia (36). Another possible cause is hyperventilation caused by anxiety (37). In normal persons, end-tidal mixed alveolar PCO_2 shows only minor fluctuations from the mean value. A single deep breath however, may lower this by 10 mm Hg or more and this situation takes 20-30 seconds to recover. Others have shown that arterial PCO_2 decreases to values between 2.4 and 3.6 kPa in 10 minutes of hyperventilation (38). Causes for this may be found in various factors such as fever (absence of post-analytical temperature correction) or non-physiologic stimuli to ventilation (medication, bacterial toxins). Finally, shifts in PCO_2 setpoint to lower values have been observed in high altitude adaptation (37), and in exercise (39,40). A similar shift in setpoint caused by illness may explain why patients tend to feel better with slight overventilation. The concept of setpoint will be discussed in more detail in the next section (5.12).
The mean values for children are somewhat, but not dramatically shifted with respect to those obtained from healthy individuals (Table 5.5 and ref. 41). The base excess is not included because this information was not available in reference 41.

5.12 Setpoint

The term setpoint is used to indicate the physiological level at which a variable is regulated. The concept of setpoint is of utmost importance for a better understanding of the processes in pathology. Examples of setpoints in physiology are those for body temperature (42), the body weight (43), PCO_2 (39), osmolality (44), and the setpoint for calcium homeostasis (45). Examples of shifts in setpoints have been published.
If a setpoint can shift from its normal value, there must be a path that is followed from the old to the new value. Furthermore, a path implies a direction in which the shift is

Table 5.5 *Means and standard deviations of reference regions for children.*

	Multivariate (patient-based)			Univariate (health-based)		
	Age	Mean	s.d.	Age	Mean	s.d.
pH	1m-1y	7.36	0.04	3m-2y	7.40	0.03
	1-5y	7.38	0.04	1-3y	7.38	0.03
	5-18y	7.39	0.04	3-15y	7.41	0.03
PCO_2 (kPa)	1m-1y	5.65	0.95	3m-2y	4.52	0.53
	1-5y	4.85	0.57	1-3y	4.52	0.53
	5-18y	5.01	0.52	3-15y	4.92	0.40

taking place. In the univariate case, the direction can be either an increase or a decrease. Well-defined examples of such paths for the multivariate case can be found in the literature describing acid-base disorders (*e.g.* ref. 32). During a shift, there is either no setpoint or a setpoint that is variable in time. If the former were true, regulation would have been lost and various paths would be possible. This is in contradiction with the existence of a well defined direction of the shift. If the latter is true, then the meaning of the term 'set' point is weakened. Indeed, small circadian fluctuations in temperature setpoint have been reported.

It would appear that a setpoint is a compromise between the influence of macroscopic physiological factors on a microscopic regulatory level, and that the path followed during its shifting is the inevitable resultant from these macroscopic factors. In this view, the setpoint is a deterministic concept of the strive for homeostasis.

It will be clear that this way of thinking exerts attraction on those who consider every patient as a unique being. Others consider a patient as an individual that is part of a group of similar individuals, who are all subject to the same laws of nature. In the former way of thinking, the patient is considered as its own reference, leading to the determination of intra-individual reference values. In the latter, reference values are determined from the group as a whole. Yet, at least for acid-base physiology, these shifts in setpoint are known and can be calculated with their relatively narrow 95% confidence regions (32). This means that these shifts are comparable to a high degree between the members of a group of patients.

In summary, there seem to be three levels of refinement: the properties of a group of individuals (reference range), of a single individual (setpoint), and of this individuals condition (shifted setpoint). The shift in setpoint appears to be a temporary surrender of nature to the conditions imposed, in order to preserve physiological integrity.

5.13 Evaluation of the multivariate model

In order to investigate the possible usefulness of the reference ellipsoid, we have incorporated the calculation of the Mahalanobis distance in the data-processing program connected to the blood-gas apparatus. On the basis of two data sets, the performance of the Mahalanobis distance in classifying abnormal blood-gas values was evaluated against the judgement of two clinicians.

From the hospital population, 100 consecutive acid-base measurements were obtained with corresponding squared Mahalanobis distances. By definition, a value of d^2 exceeding 7.8 was considered abnormal. The list of primary data (pH, PCO_2, and base excess) without other information about the patients was presented to a senior house officer, considered to be an expert in the field (indicated as '1st clinician' in Table 5.6). He was asked to mark the data sets which he considered to be abnormal.

A second data set of 100 measurements was constructed, consisting of 95 serious acid-

Table 5.6 *Evaluation of acid-base data.*
MV = multivariate. UV = triple univariate. A squared Mahalanobis distance ≤ 7.8 is considered normal, a value > 7.8 abnormal. In the univariate classification, all values simultaneously within the reference ranges 7.35–7.45 for pH, 4.7–6.4 kPa for PCO_2 and −3 to +3 meq/L for base excess are considered normal.

		First set		Second set	
		1st clinician		1st clinician	
		normal	abnormal	normal	abnormal
MV	normal	49	13	5	0
	abnormal	5	33	11	84
UV	normal	27	4	7	0
	abnormal	27	42	9	84
		2nd clinician		2nd clinician	
		normal	abnormal	normal	abnormal
MV	normal	41	21	5	0
	abnormal	4	34	30	65
UV	normal	24	7	7	0
	abnormal	21	48	28	65
		2nd clinician		2nd clinician	
		normal	abnormal	normal	abnormal
1st clinician	normal	43	11	16	0
	abnormal	2	44	19	65

base abnormalities (d^2 values in the range 8.5–760) and 5 univariate normal cases. The same clinician was asked again to evaluate the list.
In order to evaluate between-observer agreement, a second clinician from the same department was asked to evaluate both data sets in the same manner. The results of this experiment are summarized in Table 5.6.
From this Table, the kappa statistics (±s.d.) can be calculated (46):

Table 5.7 *Evaluation of agreement: kappa values and standard deviations. Abbreviations are the same as in Table 5.6. clin. = clinician.*

	First set	Second set
First clin./MV	0.63±0.08	0.43±0.16
First clin./UV	0.40±0.09	0.58±0.13
Second clin./MV	0.51±0.08	0.18±0.12
Second clin./UV	0.42±0.09	0.24±0.12
First clin./Second clin.	0.74±0.07	0.52±0.10

The kappa statistic is a measure of agreement corrected for agreement by chance. Kappa equals zero if there is agreement by chance only, and unity if there is complete agreement. Although the value depends on the composition of the test set, it is unbiased by agreement by chance.
From the kappa values in Table 5.7 it can be seen that in the first set the first clinicians judgement agrees more with the multivariate than with the univariate method ($p < 0.06$, ref.46), whereas the second clinicians judgement compares equally with both methods. This tendency to a multivariate approach of the first clinician was no longer apparent in the second set with the more serious cases. Furthermore, the second clinicians judgement seems to be influenced more than that of the first by the difference in composition between the two test sets, reflected by the much lower kappa values. The first of these (kappa = 0.18) does not reach significance compared to agreement by chance (kappa = 0), whereas the second is just significant at a p value of 0.05. The agreement among both clinicians was the best for the first set and decreased significantly for the second set ($p < 0.08$).
The relatively low measure of agreement of both clinicians with the conventional univariate method is striking. The first clinician said to weigh deviations in PCO_2 more heavily than deviations in the pH or the base excess. This seems a logical approach with artificially ventilated patients.
This evaluation is of course too limited to warrant any definitive conclusion about the value of the multivariate approach. However, a preliminary judgement can be given in that it compares reasonably well with the judgement of two experienced clinicians who were given the test results only.

5.14 The variation of the Mahalanobis distance in time

Apart from the information contained in the absolute value of the Mahalanobis distance, there is useful information contained in its change with time. This change is defined as the ratio of the new minus the old squared Mahalanobis distance and the time difference in hours, calculated from two consecutive acid-base measurements. This change, which may be negative (improvement), zero (stationary), or positive (deterioration) will be denoted hereafter as v (velocity). The value of v is strongly dependent on the time interval: samples taken with an interval of 10 minutes may generate high values of v, whereas samples taken once each day may generate v values close to zero. Considering these arguments and the sampling frequency, data sets of several patients were obtained and evaluated for time differences between 1 and 12 hours. Table 5.8 shows the frequency distribution of v values in three patients.

Table 5.8 *Frequency distribution of v values (= $\delta d^2/\delta t$).*

v [h^{-1}]	Patient 1	Patient 2	Patient 3
< −20	4	2	4
−20 - −15	1	0	1
−15 - −10	4	1	3
−10 - −5	6	1	1
−5 - 0	34	42	30
0 - 5	31	44	35
5 - 10	9	2	5
10 - 15	1	1	1
15 - 20	2	0	0
> 20	2	5	2
Total	94	98	82
Mean δt [h:min]	5:22	5:02	3:30

From Table 5.8 it can be seen that the majority of changes is found between −5 and +5 h^{-1}: 69%, 88%, and 79% of the values, respectively. The distribution is centered around zero, and approximately the same number of positive and negative changes are found. If this were not the case, the patients situation would be either predominantly deteriorate or improve. Deterioration or improvement is reflected in both the sign and in the relative magnitude of the positive and negative changes in Mahalanobis distance, respectively.
The mean sampling interval is given because relatively shorter time intervals δt (reflected in a lower mean interval) may widen the distribution (causing higher ratios and therefore more values in both tails of the distribution of v). However, this is not the case with patient 3. The initial situation of this patient had been very serious: pH 7.25, PCO_2 3.6 kPa, base excess −14 meq/L, and a squared Mahalanobis distance of 1013. This situation improved quickly on treatment. In this case, the lower mean sampling interval apparently reflects the ongoing concern for this patient, leading to frequent sampling, despite the absence of serious derangments.
We conclude from Table 5.8 and similar observations that a reference change of 5 h^{-1} appears an acceptable rough cut-off limit for signalling rapid shifts in acid-base equilibrium. Such a change could occur in spite of a normal squared Mahalanobis distance, for example from 0.2 to 6.0 in one hour (v = +5.8 h^{-1}). In such a case, an early signalling of a rapidly deranging acid-base condition is possible, which enables the clinician to take preventive measures. In combination with the absolute value of the squared Mahalanobis distance, the possibility is created to alert the physician for four kinds of conditions:

1. deviating and unstable
2. deviating but stable
3. normal but unstable
4. normal and stable

In figure 5.3 a scatter plot is presented of paired squared Mahalanobis distance and velocity values. The general impression is that high d^2 values are associated with high velocities (deviating and unstable), especially for the positive velocities.

The four classifications mentioned above may be brought to the attention of the physician by presenting the measured sets in a colour code on a terminal screen or a printed report, *e.g.* red for class 1, yellow for class 2, blue for class 3, and green for class 4. Alternatively, a department may choose to be alerted to only a subset of these types of condition. The monitoring of the Mahalanobis distance changes with time may be helpful to reduce the number of unnecessary blood-gas measurements in intensive care units.

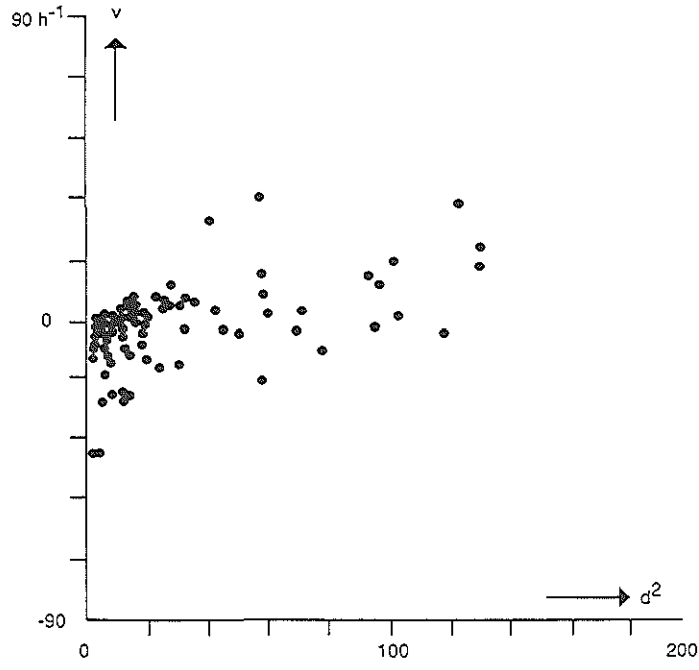


Figure 5.3 *Scatter plot of the squared Mahalanobis distance and its change in time.*

5.15 Conclusions

The use of multivariate reference ranges in clinical chemistry can have advantages for both the laboratory and the clinic. The production of such ranges from large collections of patient data is possible, but the conditions for and limitations of this method should be kept firmly in mind. At the stage of investigation described above the torch was handed over to the department of Medical Informatics of the Erasmus University Rotterdam. Further lines of investigation are directed at:

1. evaluation of the multivariate approach in a neonatal ward,
2. evaluation of the multivariate approach in another hospital, and
3. development of a multivariate approach to acid-base nomenclature.

5.16 References

1. Albert A, Harris EK
   Multivariate interpretation of laboratory data. In: Statistics, textbooks and monographs. New York: Marcel Dekker, 1987;75:1-312.

2. Iizuka Y, Kume H, Kitamura M
   Multivariate delta check method for detecting specimen mix-up. Clin Chem 1982;28:2244-8.

3. Lacher DA
   Relationships between delta checks for selected chemistry tests. Clin Chem 1990;36:2134-6.

4. Cramér H
   Mathematical methods of statistics. Morse M, Robertson HP, Tucker AW, eds. Princeton: Princeton University Press, 1946.

5. Goldschmidt HMJ
   The application of multivariate statistical analysis in clinical chemistry and haematology. [Dissertation]. Utrecht, The Netherlands: Rijksuniversiteit Utrecht, 1987. 273 pp.

6. Boyd JC, Lacher DA
   The multivariate test range: an alternative interpretation of multitest profiles. Clin Chem 1982;28:259-65.

7. Madias NE, Adrogué HJ, Horowitz GL, Cohen JJ, Schwartz WB
   A redefinition of normal acid-base equilibrium in man: carbon dioxide tension as a key determinant of normal plasma bicarbonate concentration. Kidney Int 1979;16:612-8.

8. Rehpenning W, Harm K, Domesle, Voigt KD
   Falsch positive Werte bei der Vielfachanalyse: Die Abschätzung ihrer Häufigkeit mit der Sylvesterschen Formel und ihre Reduktion durch eine multivariate Testgröße. J Clin Chem Clin Biochem 1979;17:565-72.

9. Bernhardt W, Weisner B
Bivariate wertung von Befunden: Immunoglobulin G und Albumin des Liquor Cerebrospinalis. J Clin Chem Clin Biochem 1976;14:443-7.

10. Grams RR, Johnson EA, Benson ES
Laboratory data analysis system: Section III-multivariate normality. Am J Clin Pathol 1972;58:188-200.

11. Albert A
Atypicality indices as reference values for laboratory data. Am J Clin Pathol 1981;76:421-5.

12. Schoen I, Brooks SH
Judgement based on 95% confidence limits. A statistical dilemma involving multitest screening and proficiency testing of multiple specimens. Am J Clin Pathol 1969;53:190-5.

13. Naus AJM, Borst A, Kuppens PS
Determination of n-dimensional reference ellipsoids using patient data. J Clin Chem Clin Biochem 1982;20:75-80.

14. Kågedal B, Sandström A, Tibbling G
Determination of a trivariate reference region for free thyroxine index, free triiodothyronine index, and thyrotropin from results obtained in a health survey of middle-aged women. Clin Chem 1978;24:1744-50.

15. Harris EK
Statistical aspects of reference values in clinical pathology. Prog Clin Pathol 1981;8:411-3.

16. Amador E, Bartholomew PH
Indirect methods for estimating the normal range. Am J Clin Pathol 1969;52:538-46.

17. Elveback LR, Guillier CL, Keating FR
Health, normality, and the ghost of Gauss. J Am Med Ass 1970;211:69-75.

18. Solberg HE
Statistical treatment of reference values in laboratory medicine: Testing the goodness-of-fit of an observed distribution to the Gaussian distribution. Scand J Clin Lab Invest 1986;46 (suppl 184):125-32.

19. Pittman JA, Dailey GE, Beschi RJ
Changing normal values for thyroidal radioiodine uptake. New Engl J Med 1969; 280:1431-4.

20. Leijnse B
Is Ockham's scheermes waardevol in de geneeskunde? Departing lecture at the Erasmus University Rotterdam (Dutch). 23th March 1990.

21. Naus AJM
De berekening van referentiewaarden in de klinische chemie uit analyseresultaten van een patientenpopulatie. [Dissertation]. Maastricht, The Netherlands: Rijksuniversiteit Limburg, 1982. 154 pp.

22. Owen JA, Campbell DG
A comparison of plasma electrolyte and urea values in healthy persons and in hospital patients. Clin Chim Acta 1968;22:611-8.

23. Peenen van HJ, Lindberg DAB
The limitations of laboratory quality control with reference to the "number plus" method. Am J Clin Pathol 1965;44:322-30.

24. Pryce JD
Level of haemoglobin in whole blood and red blood-cells and proposed convention for defining normality. Lancet 1960;2:333-6.

25. Merkouriou S, Dix D
Estimating reference ranges in clinical pathology: an objective approach. Statistics in Medicine 1988;7:377-85.

26. Pèpe P, Tisserand-Perrier M
Méthodes statistiques dans les sciences humaines. 1st ed. Paris: Masson et Cie., 1962.

27. Bhattacharya CG
A simple method of resolution of a distribution into Gaussian components. Biometrics 1967;23:115-35.

28. Hoeke JOO
Aspecten van de 'gemiddelden van normalen'-methode in de klinische chemie. [Dissertation]. Rotterdam, The Netherlands: Erasmus Universiteit Rotterdam, 1976. 87 pp.

29. Tietz NW
Textbook of clinical chemistry. 1st ed. Philadelphia: W.B. Saunders Company, 1986.

30. Wulkan RW, Leijnse B
Experience with expert systems in clinical chemistry. In: Kerkhof PLM, van Dieijen-Visser MP, eds. Laboratory data and patient care. New York: Plenum Press, 1988:117-24.

31. Kaldor G, Rada R
Computerised evaluation of acid-base disorders based on a nine-cell decision matrix. Med Biol Eng Comp 1985;23:269-73.

32. Pincé H, Verberckmoes R, Willems JL
Computer aided interpretation of acid-base disorders. Int J Biomed Comp 1990;25:177-92.

33. Mahalanobis PC
On the generalized distance in statistics. Proc Natl Inst Sci India 1936;2:49-56.

34. Gelsema ES, Leijnse B, Wulkan RW
Detection of aberrant observations in a background of an unknown multidimensional Gaussian distribution. Meth Inform Med 1990;29:236-42.

35. Gelsema ES, Leijnse B, Wulkan RW
A multi-dimensional analysis of three chemical quantities in the blood. Med Inform 1991;16:43-54.

36. Pootrakul P, Yansukon P, Thongmee A, Finch CA
A mild uncompensated alkalosis in anemia. Am J Med Sci 1987;294:408-11.

37. Lum LC
Hyperventilation and anxiety state. J Roy Soc Med 1981;74:1-4.

38. Arbus GS, Hebert LA, Levesque PR, Etsten BE, Schwartz WB
Characterization and clinical application of the "significance band" for acute respiratory alkalosis. New Engl J Med 1969;280:117-23.

39. Oren A, Wasserman K, Davis JA, Whipp BJ
Effect of CO_2 set point on ventilatory response to exercise. J Appl Physiol 1981;51:185-9.

40. Sharp RL, Williams DJ, Bevan L
Effects of controlled frequency breathing during exercise on blood gases and acid-base balance. Int J Sports Med 1991;12:62-5.

41. Cohen JJ, Kassirer JP
Acid-base. 1st ed. Boston: Little, Brown and Company Inc., 1982.

42. Guyton AC
Textbook of medical physiology. 8th ed. Philadelphia: W.B. Saunders Company, 1991.

43. Keesey RE
The body weight setpoint. What can you tell your patients? Postgrad Med 1988;83:114-27.

44. Rose BD
Clinical physiology of acid-base and electrolyte disorders. 2nd ed. London: McGraw-Hill Book Company, 1984.

45. Auwerx J, Demedts M, Bouillon R
Altered parathyroid set point to calcium in familial hypocalciuric hypercalcaemia. Acta Endocrinol 1984;106:215-8.

46. Schouten HJA
Statistical measurement of interobserver agreement. Analysis of agreements and disagreements between observers. [Dissertation]. Rotterdam, The Netherlands: Erasmus Universiteit Rotterdam, 1985. 131 pp.

Chapter 6

EXPLORATION OF HEMO, AN EXPERT SYSTEM FOR HEMATOLOGY

6.1 Introduction

Anemia is a frequently encountered condition in medicine. It is defined as a deficiency of circulating functional hemoglobin. The anemias can be classified on the basis of morphological characteristics (red blood cell size and color), or by physiological criteria such as erythrocyte production, erythrocyte degradation or blood loss.
Conventionally, the human eye and microscope served to investigate the red cell morphology. Technological advances have permitted the use of hemocytometry, which combines greater accuracy and precision with shorter analysis time. Instruments employing optical or impedance measurement principles produce a large amount of information from a single blood sample. Typically, the hemoglobin concentration (Hb) is measured, the number of erythrocytes per volume (RBC) as well as their size distribution, the leukocyte count, the thrombocyte count, and usually some form of leukocyte differentiation. From these primary measurements, secondary parameters such as the mean red cell volume (MCV) and the red cell volume distribution width (RDW) are calculated. The latter is defined as the ratio of the standard deviation and the mean value of the erythrocyte volume distribution curve, expressed as a percentage.

The subject of this chapter is to investigate the development of an expert system that supports the diagnostic process by automated interpretation of laboratory data and contributes to accelerated laboratory testing.
Such a system will be of value in particular in interpreting results from patients in the diagnostic phase, usually at the time they make their first, policlinical visit to the hospital. Hemocytometrical data are commonly requested in these cases. The patient goes home after such a visit, and is generally requested to return after several weeks for follow-up. If necessary, further hematological parameters will be requested, evaluated, and so on. The evaluation of hospitalized patients is more complicated because it requires clinical information about recent surgery, transfusion history, and about the application of medication influencing hematological parameters.
In the policlinical application, the automated data interpretation may result in a gain in efficiency, time, use of sample, and reduction in cost. The expert system should ideally

generate advice for follow-up measurements. If fiatted by the physician in a quick response, the laboratory is enabled to use the remainder of the sample for such additional measurements. Thus, all measured parameters are representative of the situation at one moment, instead of being separated by a large period in time (which may necessitate repeated measurements). This approach may also result in a reduction of the number of policlinical visits for each patient, since the physician will have more diagnostic information at the next visit, which provides him with a better defined picture of the diagnostic problem.

6.1.1 Physician performance

In order to appreciate the need for a diagnostic support system, a literature evaluation of the quality of diagnostics is given below:

1. In a study of 199 anemic patients admitted to the medical service of a university hospital, the recognition rate of anemia was examined as a function of the hemoglobin concentration (1). Only at levels below 5.5 mmol Fe^{2+}/L did the recognition rate reach 100%. A quarter of all anemic patients was not recognized as such, and a fifth was recognized but either misidentified or not evaluated. Some medical teams were reported to identify most of the cases of anemia, whereas others rarely did.
2. In a study of 728 anemic patients, 259 were judged to need diagnostic evaluation of the anemia (2). Of these, a fifth had diagnostic evaluations that were considered inadequate. The personal performance of the 26 medical interns in this study was expressed as the percentage of adequately evaluated patients, and varied from 25% to about 90%. The most common shortcoming of the incomplete evaluations was a failure to evaluate possible iron deficiency in normocytic patients.
3. In a study of 30 cases of anemia, the mean percentage of agreement among six expert hematologists in grading their diagnoses into four classes of probability was only 48% (3). The individual grading of a diagnosis was called satisfactory if the other experts agreed, or if it was considered a reasonable alternative. This individual percentage of rating varied from 60% to 100%. With an expert system called ANEMIA, this percentage was 83% (3).
4. In a study of 173 cases of anemia, two specialists in internal medicine who recently had followed a course in hematology, made the correct diagnosis in only 32% and 30% of the cases. However, the correct diagnosis was among the five diagnoses considered most likely in 68% and 77% of the cases (4). Corresponding figures for the expert system 'ADH' were 53% (first) and 87% (among five) respectively.
5. In another study, 7 cases of microcytic red cell disorders were presented to 17 expert physicians. Only 1 physician had the correct diagnosis in every case (5). All physicians correctly diagnosed at least 4 cases.

6. In a study of 53 cases with microcytosis and RBC > 5 × 10^{12}/L, in 41% of the cases thalassemia was not considered in the differential diagnosis by residents in internal medicine (6). A correct interpretation of the hemoglobin electrophoresis results was given in only 56% of the cases. Iron therapy was given in 17% of the cases.

These literature studies clearly demonstrate that there is sufficient reason for the development of systems that offer diagnostic assistance in anemia.

6.2 Analysis of the diagnostic process

For the development of an expert system in hematology, the diagnostic value of generally applied tests was investigated and an analysis of request patterns was made.

6.2.1 Selection of diagnostic parameters

Definitions that will be used in this section are:

Sensitivity
: The percentage of the patients with a disease having an *abnormal* test result (true positive fraction × 100).

Specificity
: The percentage of patients without the disease having a *normal* test result (true negative fraction × 100).

Positive predictive value
: The patients *with* the disease and an *abnormal* test result, expressed as a percentage of *all* patients with an *abnormal* test result.

Negative predictive value
: The patients *without* the disease and with a *normal* test result, expressed as a percentage of *all* patients with a *normal* test result.

The introduction of hemocytometry led to the availability of several parameters describing characteristics of the population of erythrocytes. By definition, the MCV is used to differentiate between microcytic, normocytic, and macrocytic anemia. The RDW, an objective and precise measure of anisocytosis (7), was claimed to differentiate between iron deficiency and heterozygous thalassemia. Finally, the reticulocyte count is indispensable in distinguishing hyper-proliferative from hypo-proliferative anemias.
Several authors have published diagnostic findings in correlation with MCV and RDW, with sometimes conflicting results (8-12).
The interpretation of these studies requires some caution (13). The factors influencing the prevalence of a disease in the selected population, should be carefully considered. The selection of patients in a study is of importance, *e.g.* hospitalized patients, policlinical

patients, 'healthy' subjects being screened for disease or a mixture of these. Ethnic and racial composition should be known, as well as age and sex distribution. Some authors present results obtained from unselected hospitalized patients, others have selected only patients with anemia, and still others selected patients with microcytosis regardless of the presence of anemia. Also, the co-occurrence of other diseases known to affect the screening measurements should be considered. The choice of the analytical instrument and methodology may also affect the results: the calculation of the RDW in the early types of cell counting instruments differs from the way it is done at present (14). Some authors use different methods of demonstrating the presence of a disease (*e.g.* in case of iron deficiency: the serum ferritin concentration, bone marrow iron staining, transferrin saturation or iron therapy responsiveness). The choice of the decision threshold to distinguish between a normal and an abnormal test result has a major influence on the results.

6.2.1.1 The mean erythrocyte volume and red cell distribution width

The following studies give an impression of the diagnostic utility of these parameters.

1. MCV in screening for nutritional deficiencies
In a series of 69 iron deficient patients, only 61% were identified on the basis of a low MCV (15). A low mean corpuscular hemoglobin concentration performed slightly better (74%). In the latter case, the positive and negative predictive values were 65% and 80%, respectively. In the same study, 89 patients were vitamin B_{12} deficient, of which only 48% could have been detected by a high MCV. The positive and negative predictive values were 32% and 93%, respectively. Also, 38 patients were folate deficient, of which 37% would have been detected by a high MCV. For this test, the positive and negative predictive values were 25% and 89%. No data were included about the occurrence of other causes for macrocytosis in this Irish population.

2. MCV and RDW in iron deficiency
In another study, a sensitivity of 56% was found for a low MCV as a screening parameter for iron deficiency anemia (prevalence 8%) in screening a population of Japanese female students (16). The claimed specificity of 96%, however, does not take the 'borderline' cases (prelatent and latent iron deficiency) into account as 'absence of disease', and thus the true specificity is expected to be lower. In the same study, the RDW had a sensitivity of 77% and a specificity of 83% for iron deficiency. Unfortunately, no information was given about the prevalence of thalassemia or hemoglobinopathy in this population.

3. RDW in microcytosis
In a study of patients with microcytosis, 94% of the patients with iron deficiency had an

elevated RDW (11), which was also seen in 48% of the patients with thalassemia. The prevalence of thalassemia in this Hawaiian population with microcytosis was high (53%) compared to that of iron deficiency (37%). The differentiation between these disorders is important because the thalassemics are at risk of receiving unnecessary and even potentially harmful iron therapy.
The RDW has been shown to correlate positively with the reticulocyte count in a group of 1121 Jamaican individuals with homozygous sickle cell disease, but not in 344 individuals with hemoglobin SC disease (17). Almost all patients in the former group had an elevated RDW.

4. MCV and RDW in iron deficiency and heterozygous β-thalassemia
In a thorough Italian study of outpatients with microcytosis, 33% of the patients with iron deficiency had RDWs which were in the range for β-thalassemia trait (18). Only 28% of the thalassemia patients had RDWs within the normal reference limits, the remainder had higher RDWs. Of the thalassemia group, 23% were not anemic. The authors presume that the thalassemia patients with more severe anemia tend to have higher RDW. This has also been presumed elsewhere (8).

5. RDW in anemia of chronic disease and in thalassemia trait
In a Californian study, an increased RDW was found in 66% of the patients with heterozygous β-thalassemia, in 32% of the patients with anemia of chronic disease, in 34% and 53% of the female and male patients, respectively, with α-thalassemia trait (19).

A review of the literature on RDW values in disorders which may present with microcytosis resulted in the information presented in Table 6.1.
The conclusion from these data is that the ranges of RDW for these conditions show wide overlap and do not permit a differentiation between these disorders, in particular not in individual cases. Only very high RDWs (*e.g.* >25, from Table 6.1) associated with microcytosis may be attributed to iron deficiency, and only if other conditions causing such high values have been excluded (homozygous sickle cell disease, sickle cell β°-thalassemia, HbSC, or HbH disease).
Reconsidering the definition of the RDW, it is apparent that any mixture of two erythrocyte populations with sufficiently different mean red cell volumes will increase the RDW. Such a mixture may be composed of normocytic and microcytic (iron deficiency, chronic inflammation), normocytic and macrocytic (vitamin B_{12} deficiency), or microcytic and macrocytic cells (combined iron and folate deficiency). Since the RDW is the ratio of the standard deviation of the erythrocyte volume distribution curve and the MCV, its value will be influenced by both measurements. It is therefore not a pure measure of anisocytosis and the diagnostic value of this parameter is largely diminished.
Another parameter has been proposed as a measure for red cell volume dispersion. This

Table 6.1 *The red cell distribution width (RDW) in three disorders, which can produce microcytosis. The number of patients (n) and the range of possible values are given.*

Literature reference	Iron deficiency		β-Thalassemia minor		Inflammation	
	n	range	n	range	n	range
8[14,1]	67	(12.7-19.9)	16	(10.5-16.9)	28	(12.8-18.8)
20[1]	43	(11.3-29.9)	28	(12.9-18.6)		
21[15,2]	69	(17.2-33.5)[3]	67	(12.8-20.5)[11]		
22[14,1]	116	(10.0-36.7)[4]	186	(10.8-22.5)[11]		
18[15,2]	196	(13.0-33.8)[5]	358	(12.5-21.7)[11]		
23[14,1]	263	(11.5-23.9)[6]	445	(13.1-17.2)[11]		
11[15,2]	68	(13.5-29.4)[5]				
16[17,2]	43	(12.2-19.9)[7]				
24[16,2]	17	(16.3-27.7)[8]			17	(14.5-18.4)[12]
10[15,2]	10	(16.8-25.0)[9]			13	(12.1-25.4)
25[15,2]	76	(12.9-23.3)[10]			22	(12.9-20.2)[13]
This thesis[15,2]	104	(12.9-33.1)				

1. Mean ± 2s.d.
2. Limit values
3. Diagnosed by serum ferritin, transferrin, iron, erythrocyte protoporphyrin, and responsiveness to iron
4. Diagnosed by serum iron and transferrin
5. Diagnosed by serum iron, transferrin, and transferrin saturation
6. Diagnosed by serum ferritin and iron-responsiveness
7. Diagnosed by serum ferritin and transferrin saturation
8. Diagnosed by bone marrow iron or by serum iron, transferrin saturation, and iron-responsiveness
9. Diagnosed by serum ferritin
10. Diagnosed by bone marrow iron or iron responsiveness
11. Diagnosed by HbA_2 (normal iron studies)
12. Unresponsive to iron
13. Normal ferritin
14. Coulter S Plus II
15. Coulter S Plus IV
16. Coulter C-1000
17. Sysmex E-4000.

is the absolute distribution width at half maximum ($ADW_{0.5}$). Using an H6000/H601 blood cell analyzer from Technicon Instruments, a specificity of 92% and a sensitivity of 66% were obtained in screening a hospital population for α- and β-thalassemia trait (26).

6.2.1.2 The reticulocyte count

The reticulocyte count is an indicator of erythropoetic activity. It indicates whether the expected response to anemia, *i.e.* an increased erythropoesis, has or has not occurred. Absence of this response points at bone marrow failure as a cause of the anemia. In contrast, if the reticulocyte count is increased, the cause of the anemia is usually hemolysis or blood loss. This blood loss may be obvious, *e.g.* excessive menstrual bleeding, or it may not. In the latter case, *e.g.* a gastrointestinal blood loss may be reveiled.

Therefore, an accurate reticulocyte count is of value in the recognition of these patients. The number of reticulocytes is expressed either relatively, as a percentage of the erythrocyte count, or absolute, as the number of reticulocytes per unit volume. Studies with radioactive iron have shown that with decreasing hematocrit, the residence time of these cells in the bone marrow is shortened and hence the number of peripheral reticulocytes increases (27). The reticulocyte count can be corrected for this shift in residence time and for the hematocrit value, resulting in a reticulocyte production index (27). Another refinement can be made by classifying the reticulocytes according to their RNA content, which decreases with cell age (28).
The manual counting of reticulocytes is tedious and subject to a large imprecision. Inter-laboratory variations of 25-48% have been found in the U.S.A. and in France, and similar intra-laboratory variation coefficients have been demonstrated (28-30). We have determined an analytical intra-individual variation coefficient of 18% (at a mean value of 3.3% reticulocytes) and an analytical inter-individual variation coefficient of 36% (from duplicate measurements).
With the equipment that has been developed for automated reticulocyte counting, variation coefficients between 4% and 8% can be attained (28,31). These instruments provide the clinician with a more precise parameter to measure the capacity of the bone marrow to respond to the anemia. The important role of this parameter in the diagnosis of anemia is a strong argument for the introduction of a reticulocyte counter in the clinical chemistry laboratory.

6.2.2 Discriminant functions

The realization that one parameter is insufficient to distinguish iron deficiency from heterozygous β-thalassemia, has led many to try combinations of parameters. It is important to realize that a discrimination between these two disorders is something different than screening a population to detect, *e.g.* β-thalassemia minor. Most of the attention in the literature has been focused on the first problem.
A discriminant function is a mathematical expression, in which several measurement parameters may be combined. For the problem described above, combinations of Hb, MCV, RDW, RBC, MCH (mean corpuscular hemoglobin), MCHC (mean corpuscular hemoglobin concentration), MPV (mean platelet volume), and PDW (platelet distribution width) have been proposed. In any given case, the value of such an expression in relation to a decision level determines which disorder is present. Thus, a result is offered which is dichotomous by nature (present or absent). This approach lacks flexibility, especially in the area where the respective probabilities are not very different. Also, the possibility of a combination of the two disorders is ignored. These two problems can be overcome by the use of a continuous scale of probability, such as is incorporated in an expert system shell in the form of certainty factors. Discriminant functions may be used to

calculate the probability of an intermediate conclusion in the reasoning process (32).
The performance of discriminant functions in separating iron deficiency from β-thalassemia minor has been evaluated (26,32,33). One publication presents a discriminant function to screen a 'healthy' population for heterozygous thalassemia (34). This function produces very few false positive and false negative results compared to five other existing functions. Unfortunately, few cases of iron deficiency were included in the evaluation data set, so its validity in screening mixed populations has yet to be proven.
The lack of transferability of discriminant functions between analytical instruments from different manufacturers has been discussed (32).

6.2.3 Analysis of the request pattern

The patterns of requests of hematological parameters for policlinical patients with anemia were evaluated. To this end, a group of 105 anemic patients (39% men and 61% women) who had not been visiting the hospital for six months was investigated. The reference values used in this evaluation are the following:

	♂	♀
Hb (mmol Fe^{2+}/L)	8.7-10.5	7.5-9.7
MCV (fL)	82-100	
RDW (%)	11.8-15.6	
Reticulocytes (%)	< 2.0	
Reticulocyte index	2.0-3.0	
Reticulocytes (10^9/L)	9.2-114	7.8-106

The Hb, MCV and RDW were measured on a Coulter S plus IV cell counter, the reticulocytes were counted manually and expressed as a percentage of the erythrocyte count. Table 6.2 lists the frequency of the measurements.

Table 6.2 shows that the measurement of Hb is almost always requested in this group. Therefore all measurements done by the cell counter are potentially available for use in an interpretative system. Traditionally the combination of Hb and Ht (hematocrit) is requested (91% in this group). However, the Ht is a derived parameter which provides little extra information over the Hb. Other frequently requested combinations were: Hb with Ht and leukocytes (67%), and LD (lactate dehydrogenase) with bilirubin (31%).
The relative lack of interest in the MCV and RDW is remarkable. Nevertheless, 24 patients group did have an abnormal MCV, and 18 patients an abnormal RDW. Apparently the physicians hardly use these measurements in screening for disease. Also, the excessively low number of requests for ferritin, vitamin B_{12} and folic acid is striking. It might be assumed that these investigations have already been done by the family

Table 6.2 *Requests (in %) for hematology evaluation in a group of 105 polyclinical patients with anemia.*

Test	%	Test	%
Hemoglobin	94	Ferritin	5
Hematocrit	92	Transferrin	5
Leukocyte count	73	Iron	4
Bilirubin	47	Lymphocytes	4
LD	41	Monocytes	4
Thrombocyte count	38	Granulocytes	4
Erythrocyte count	15	Vitamin B_{12}	3
MCV	14	Folic acid	3
RDW	13	Reticulocytes	1

physician or that these requests are not considered to be necessary before the cell count results are available.

In order to test this hypothesis, follow-up measurements of either iron, transferrin, ferritin or vitamin B_{12} and folic acid in these patients were recorded over a period of 2 years. Iron, transferrin, or ferritin were measured in only 11% of the patients, and vitamin B_{12}, folic acid, or both in 14%. Of the follow-up measurements, 24% were done within 7 days, 28% in 7-50 days, and 48% after periods longer than 50 days. Although retrospectively it is impossible to determine if these measurements would have been necessary, the usefulness of accelerated laboratory testing is emphasized.

A request of LD and bilirubin to detect hemolysis is obvious, although, LD may also be used to detect a tumor or myocardial infarction. The most frequent combination of hematology requests and requests for other measurements was LD, bilirubin, alkaline phosphatase, and gamma-glutamyl transferase (GGT; 41%), apparently to rule out disorders of the liver as a cause of a high bilirubin concentration. Of the 47 patients with bilirubin measurements, only 2 patients had an increased value (> 14 μmol/L). Some physicians (20%) left out GGT, possibly because they were not screening for alcohol abuse.

In the groups mentioned above, the partitioning of the MCV and RDW results was investigated. The results are given in Table 6.3.

A. The policlinical group

The results of this relatively small group may not be representative for a large number of patients, however they do give an impression of the situation. The high frequency of anemias with a normal MCV and a normal RDW is striking. Within this subgroup, 30% of the patients had an increased leukocyte count, which may point to anemia of inflammation or infection.

The frequency of a decreased MCV is much higher than that of an increased MCV. This

Table 6.3 *Distribution of MCV and RDW results (in %) of three groups: A. 105 policlinical patients with anemia, B. 75 hospitalized patients with anemia and reticulocyte requests, C. 1929 randomly selected patients (hospitalized and policlinical) with anemia.*

MCV	RDW	A.	B.	C.
	Low	0	0	0
Low	Normal	10	1	5
	High	9	11	10
	Low	0	0	<1
Normal	Normal	69	56	50
	High	8	23	26
	Low	0	0	0
High	Normal	4	0	4
	High	0	9	5
Low	any	19	12	15
Normal	any	77	79	76
High	any	4	9	9
any	Low	0	0	0
any	Normal	83	57	59
any	High	17	43	41

is in line with results from another hospital in the Netherlands, where 47% of the patients had microcytosis and only 3% had macrocytosis in a population of clinical and policlinical patients with anemia.

B. and C. The hospitalized group and the random group

Because the reticulocyte count was hardly requested in the policlinical group, hospitalized patients with requests for a reticulocyte count were selected. Of the 102 patients in this group, 75 were anemic (of these, 84% were women and 16% were men). A group of 1929 hospitalized and policlinical patients with anemia and with requests for cell counting was taken for comparison. This group consisted of 40% women and 60% men.

Of the anemic patients with reticulocyte requests, 79% had a normal MCV and 57% had a normal RDW, which is almost the same as in the large random group. As many as 83% of the anemic patients with reticulocyte requests had normal reticulocyte counts. Calculation of the reticulocyte production index revealed that only 1% of the patients had an increased index, 1% had a normal index, and 98% had a decreased index. This distribution of results was exactly reproduced when prospective reticulocyte counting was done in a group of 100 hospitalized patients with Hb concentrations below 6.4 mmol Fe^{2+}/L.

A difference in distribution of RDW between the group of hospitalized patients and the policlinical group is expected. Adequate therapy in the hospital may cause an increase in the reticulocyte count, which leads to a new population of normocytic cells and thereby to an increase in RDW. The percentage of increased RDW is indeed higher in the

hospitalized group than in the policlinical group (43% and 17%, respectively).
On further analysis of the large random group, a difference according to sex became apparent. The frequency of microcytosis was higher in the subgroup of women (24%) than in the subgroup of men (9%). The same sex-related difference was also observed in the policlinical group: 27% of the women had microcytosis compared to 13% of the men. A possible cause for this difference could be an increased loss of blood by menstrual bleeding with iron deficiency as a result. In the large random group, the percentage of high RDW was essentially the same for both sexes (39% for men and 43% for women).

6.3 Anemia of inflammation

Inflammation is a frequent cause of anemia. In approximately 75% of the patients with this type of anemia the MCV is normal and in 25% it is decreased. An increased haptoglobin concentration and a transferrin concentration in the low normal range are signs of an acute phase reaction and can improve the detection rate of this anemia. Because ferritin is a positive acute phase protein, it can only with caution be used as an indicator of iron stores.
In a group of 100 patients with transferrin <62 μmol/L and ferritin > 50 μg/L, a decreased folate concentration (< 7.1 nmol/L) was found with a remarkably high frequency (50%). In a subgroup of 39 patients satisfying the above conditions and with serum iron <10 μmol/L (inflammation), this frequency was essentially the same (54%). The reason for this high frequency is unknown. In the reverse situation, 54 patients with decreased folate had the above constellation of transferrin and ferritin values in 31% of the cases.
A group of 48 patients with rheumatic arthritis (satisfying the ARA criteria) was taken to study anemia of inflammation. This group was subdivided in a group with iron deficiency (grade 1 or less stainable bone marrow iron), and a group without iron deficiency (more than grade 1 bone marrow iron). Table 6.4 shows the characteristics of both groups.

In this group, the frequency of a decreased folate concentration (23%) is lower than found in the group of 100 patients mentioned above (50%). An explanation could be the presence of diseases with a high rate of cell turnover, or folate antagonist chemotherapy which depresses the folate concentration. Indeed, in the large group, 66% of the patients had a ferritin concentration exceeding 300 μg/L, which may be caused by malignancies. However, the frequency of a decreased folate concentration in this subgroup was 47%, which is not significantly different from that in the unselected group (50%). Therefore this effect must be due to a common characteristic within this group, presumably some effect of the inflammatory reaction.

Table 6.4 *Characteristics of patients with anemia of inflammation (Hb < 7.4 mmol Fe^{2+}/L).*

	Iron sufficient	Iron deficient
Number of patients	26	22
Women/men	19/7	17/5
MCV <80 fL	1	12
80-96 fL	24	8
>96 fL	2	2
MCH <1600 amol	11	15
1600-2200 amol	15	7
>2200 amol	0	0
Iron <11μmol/L	25	22
Transferrin <62 or >62μmol/l	26/0	15/7
Ferritin <50 or >50 μg/L	3/21	18/2
Folate <7 nmol/L	6	5
Vitamin B_{12} <150 pmol/L	5	5
Reti.prod.index <2 or 2-3	22/2	18/2

6.4 Development of the expert system

When hematological disorders are considered, the hemoglobin concentration is almost always the first parameter to be evaluated. In the expert system HEMO, three domains have been set up: one for Hb below the reference range (ANEMIA), one for Hb above the reference range (POLYHEMIA), and one for Hb in the reference range (NORMOHEMIA). Various reference ranges, depending on sex and age were implemented. Because of the high frequency of anemia, efforts have been mainly directed at the development of this domain. The most practical first step in the diagnosis of anemia is the differentiation on the basis of MCV. Although its discriminating value is limited (section 6.2.1.1), it is a reasonable choice. The next step is the measurement of the reticulocyte count, to provide information about the ability of the bone marrow to respond to the existing Hb deficiency. In Table 6.5 the structure of the domain ANEMIA is given.

The disorders in Table 6.5 must be confirmed by appropriate tests, such as serum ferritine, and serum iron for iron deficiency, vitamin B_{12} or folic acid for the corresponding deficiencies, LD and bilirubin or haptoglobin for hemolysis, hemoglobin electrophoresis for thalassemia, bone marrow examination for sideroblastic anemia or bone marrow insufficiency. Starting with a complete blood count and a reticulocyte count, HEMO indicates which tests seem most appropriate to perform next.

6.4.1 Subdomain Chronic hemolysis or blood loss

The conclusion chronic hemolysis or blood loss is completely dependent on an elevated

Table 6.5 *Structure of the subcontext* ANEMIA *(decreased Hb).*

MCV	Reticulocyte count	Interpretation
increased	normal/decreased	Vit.B_{12} deficiency Folate deficiency Bone marrow insufficiency Sideroblastic anemia
normal	normal/decreased	Iron and folate deficiency Acute hemolysis/blood loss Bone marrow insufficiency Sideroblastic anemia
decreased	normal/decreased	Iron deficiency Thalassemia Inflammation
any value	increased	Chronic hemolysis/blood loss

absolute reticulocyte count and independent of the MCV. This section is subdivided on the basis of additional tests such as haptoglobin, lactate dehydrogenase, and total bilirubin (figure 6.1).

The conclusion hemolysis is made if either haptoglobin is decreased (regardless of the value of LD and bilirubin), or if LD and bilirubin are increased. The conclusion is blood loss if haptoglobin is normal or increased, and both LD and bilirubin are normal or decreased. In the case of blood loss, further examination should include a search for gastro-intestinal and gynaecological causes. We have not further extended this section.
In the case of hemolysis, the direct Coombs test can indicate if there is immune-hemolysis. The temperature sensitivity of a positive Coombs test allows differentiation between warm antibodies, cold autoantibodies, and the presence of allo-antibodies.
If the Coombs test is negative, an osmotic resistance test may detect a red cell membrane defect. If this test is normal, an acid Ham test is done to detect a paroxysmal nocturnal hemoglobinuria. If this is absent, a test is done to detect a deficiency of one of the erythrocytic enzymes. A diagnostic system for erythrocyte enzyme tests has been reported (35). If the erythrocyte enzymes have normal activity, the remaining possibilities are a hemoglobinopathy, or an anemia caused by hepatic disease. The former is confirmed by hemoglobin electrophoresis, and the latter will have become apparent at a much earlier stage from the hepatic enzymes.

6.5 Evaluation of HEMO

The best way of evaluating a diagnostic support system is by prospective testing of its performance with a large number of cases, obtained from the population for which the

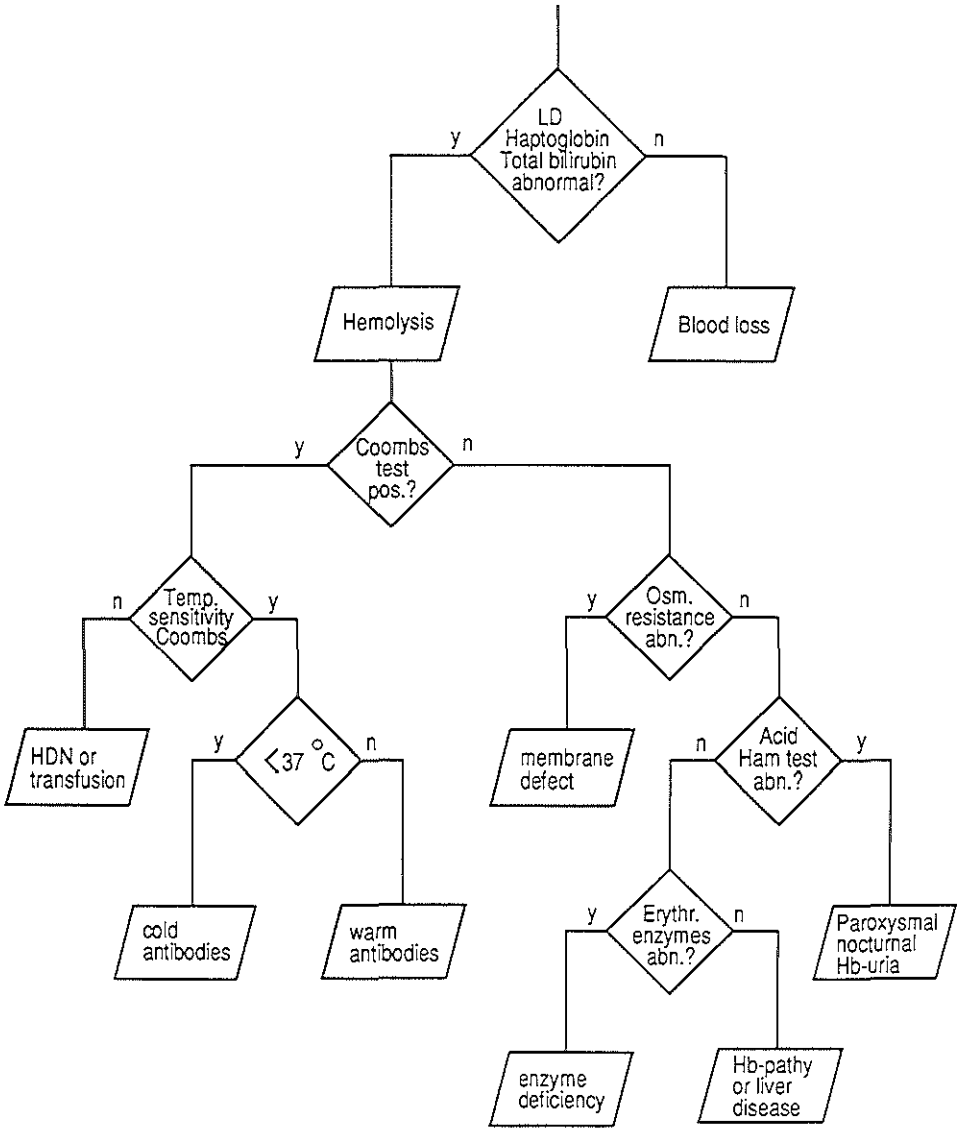


Figure 6.1 *Structure of the section Hemolysis/blood loss.*
Abbreviations are: LD = lactate dehydrogenase activity, HDN = hemolytic disease of the newborn, PNH = paroxysmal nocturnal hemoglobinuria.

system is developed. This requires the availability of laboratory results and corresponding diagnoses. Because such data were not available prospectively, the expert system HEMO was tested with retrospectively collected cases of anemia from the hospital population. For such cases, the identification number, date of diagnosis and diagnosis were obtained from the hospital registration system. The appropriate laboratory data, if present, were obtained from the hospital information system. The combination of the Hb, MCV, RBC, and relative reticulocyte count is required for the use of HEMO. This combination was found in 115 cases which are shown in Table 6.7. Each of these cases was evaluated by HEMO.

Table 6.7 *Composition of the evaluation set and performance of HEMO. For explanation see text.*

	Total	Correct conclusion	No final conclusion[1]
Iron deficiency	61	36	12
Aplastic anemia	25	8	10
Folic acid deficiency	10	2	3
Auto-immune hemolysis	10	6	1
Vitamin B_{12} deficiency	3	2	0
Thalassemia	3	1	0
Sickle cell trait	2	0	2
Spherocytosis	1	1	0

[1]Correct conclusion still included

1. Iron deficiency caused by chronic blood loss (61 cases)

There were 61 cases in this group, none of which satisfied the criteria for anemia of inflammation: serum ferritin >50 μg/L, serum transferrin <62.7 μmol/L (36). In 12 cases, HEMO did not reach an unambiguous conclusion, but in all of these cases chronic blood loss was among the diagnoses still considered likely at the end of the consultation. In the other 49 cases, shown in Table 6.8, an unambiguous conclusion was reached.

Eight cases were diagnosed as bone marrow insufficiency. All of these had a normal MCV and a normal absolute reticulocyte count in spite of the anemia, a normal or increased serum ferritin and a normal or increased serum iron. Transferrin saturation was 13% or higher in 7 cases and unknown in one case. Iron therapy as a cause seems unlikely, because all of these patients did already have this combination of iron studies at earlier measurements. Therefore the diagnosis iron deficiency seems hardly valid in these cases. HEMO could have concluded blood loss in five of these cases, hemolysis in one case, and in two cases no diagnosis was possible (unknown haptoglobin). These potential diagnoses were not reached because the diagnosis of bone marrow insufficiency was made earlier and therefore the consultation had ended. This is a serious defect in HEMO which will be discussed further. In 4 cases, no diagnosis was given because the

Table 6.8 *Conclusions made by HEMO in 49 cases of iron deficiency caused by chronic blood loss.*

Chronic blood loss	27
Bone marrow insufficiency	8
No diagnosis	4
Iron deficiency	4
Iron and folate deficiency	3
Iron deficiency and hemolysis	2
Hemolysis	1

condition of an isolated increased bilirubin concentration (normal LD and haptoglobin) was not included in the system.

In 8 cases there was an accompanying decreased folate concentration. In three cases this was recognized, in five cases it was not: in two, the diagnosis iron deficiency was made earlier and the consultation had ended. In two cases the reticulocyte count was increased, which was not included in the condition for combined iron and folate deficiency, and neither was a normal iron concentration (the last case). One patient with a coinciding decreased vitamin B_{12} concentration was classified as having bone marrow insufficiency (normal MCV).

2. Aplastic anemia (25 cases)

In 8 cases the correct conclusion (bone marrow insufficiency) was made, in 4 chronic blood loss, in 1 iron deficiency (low iron and ferritin), and in 1 iron and folate deficiency (based on low iron and low folate). In 10 cases an unambiguous diagnosis was not reached, but the correct conclusion was included in the final set. In the aplastic anemia cases, 73% of the patients had a decreased hemoglobin concentration, a decreased leukocyte count ($<4.0 \times 10^9$/L) and a thrombocyte count less than 200×10^9/L. The majority of the remaining 27% of patients had received a recent blood transfusion. In 1 case there was no final conclusion, but the correct conclusion had been ruled out.

3. Folic acid deficiency (10 cases)

In 1 case the diagnosis was folate deficiency, in 1 case iron and folate deficiency, in 5 cases bone marrow insufficiency. In two of these five cases the MCV was normal, which was not in the condition for folate deficiency, and in three bone marrow insufficiency was concluded earlier and the consultation had ended. In 2 out of 3 other cases the correct diagnosis was included in the final set. In one it was excluded because the increased vitamin B_{12} concentration was not in the condition for folate deficiency.

4. Auto-immune hemolysis (10 cases)

In 5 cases the diagnosis was immune hemolysis, in 1 case iron and folate deficiency (based on low iron and folate) and hemolysis, and in 3 cases bone marrow insufficiency. In one of these three the cause was an isolated increased LD (normal bilirubin, unknown haptoglobin) which was not provided as a condition for hemolysis. In two, bone marrow

insufficiency was concluded, before the result of the Coombs test could have been requested; however, an intermediate conclusion of hemolysis was added. In 1 case, the correct conclusion was in the final set.
5. Vitamin B_{12} deficiency (3 cases)
In 1 case vitamin B_{12} deficiency was concluded, in 1 case vitamin B_{12} and folate deficiency, and in 1 case bone marrow insufficiency (B_{12} deficiency not detected because of a normal MCV).
6. Thalassemia (3 cases)
In 1 case thalassemia was concluded. Two other cases were not identified because there was no hemolysis, which was in the condition for thalassemia.
7. Sickle cell trait (2 cases)
In the 2 cases of sickle cell trait, the correct diagnosis was included in the final set.
8. Spherocytosis (1 case)
The case of spherocytosis was correctly detected as an erythrocyte membrane defect.

This evaluation of HEMO disclosed several imperfections.
The most important of these is the fact that HEMO accomodates only one diagnosis for each case and will stop as soon as one has been found. Since not all diagnoses are mutually exclusive, some of them may be missed. Several combinations (iron and folate deficiency, iron deficiency and hemolysis) have been incorporated in separate rules. When the measurements satisfy more than one diagnosis (*e.g.* bone marrow insufficiency and folate deficiency), the outcome becomes dependent on which diagnosis is encountered first. This serious drawback can be overcome in several ways.
The utility factor of the rules concluding the infrequently occurring anemias can be changed to a lower value. As a consequence these rules will be consulted only after all other rules have been tried. A better solution to this problem is to enable the system to accomodate several diagnoses at a time. Another possibility is to alter the affirmative approach to an exclusive approach: the system then reaches its final conclusion(s) by exclusion of all solutions that are incompatible with the measurements. Several minor imperfections can be classified as too tight conditions: *e.g.* folate deficiency is concluded only if the vitamin B_{12} concentration is normal. Preceding therapy is likely to result in a raised vitamin B_{12} level, so this should be included in the condition.
Also, the possibility of a low folate and unknown vitamin B_{12} concentration should be included. Conditions with a normal haptoglobin concentration should also include an increased concentration, because of its behaviour as a positive acute phase protein. Conditions of isolated increases in LD or bilirubin should be incorporated.

In many cases data were partially absent. The most important of these was the reticulocyte count (absent in 60% of 150 cases of iron deficiency by chronic blood loss). In the same group, the ferritin concentration was missing in 37% of the cases, and in 49%

the combination of iron and transferrin. Vitamin B_{12} and/or folate concentrations were measured in 53% (80 cases), half of which were normocytic, and half microcytic. This led to the discovery of 16 cases with low folate, and 4 with low vitamin B_{12} concentration. Haptoglobin was measured in 31% of the cases. This parameter was not strictly needed by HEMO because LD and bilirubin were measured in almost all cases.

6.6 Systems for automated interpretation of hematological disorders

In recent years, large efforts have been made to construct interpretative systems for anemia. Table 6.9 presents an overview of systems for the automated interpretation of hematological disorders, mostly anemia. Systems for the diagnosis of leukemia, complement diseases and transfusion reactions have been excluded (For these, see chapter 1, Table 1.1).

Table 6.9 *Systems for the automated interpretation of hematological disorders. Abbreviations: n.r. not referenced, comp. compatible*

Name	Programming language/shell	Hardware	Remarks	Literature reference
HEME	n.r.	IBM-n.r.	1	37,38
KIRON	FORTRAN	Honeywell 6040		39
ANEMIA, MICROANEMIA	FORTRAN/EXPERT	VAX 11/780, IBM-XT		3,40
NEOANEMIA	Common LISP/KEE	TI-EXPLORER		41
n.r.	n.r.	n.r.		42
ADH	n.r.	IBM-XT	2	4
n.r.	n.r./CSGP	IBM-PC		43
Consult-I	CLS	IBM-AT	3	44,45
BCDE	BASIC	IBM-PC comp.	4	46
BCDE2	BASIC	IBM-PC comp.	5	5
RBC	BASICA	IBM-XT		47
Micro Hema Screen	BASIC	Apple IIe	6	23
n.r.	n.r./MacSMARTS	Apple Mac. SE	7	48
n.r.	n.r./VP-EXPERT	IBM-PC	8,9	49

1. Copyright International Business Machines Corporation 1975
2. Aide à la Décision en Hématologie, Copyright Coultronics SA, France
3. Consult Learning System, registered trade mark of Patrick Consult Inc., Cincinnatti, Ohio
4. Blood Count and Differential Evaluation, registered trade mark of Island Analytics, Galveston, Texas 1986
5. Copyright Lea and Febiger 1988
6. On-line coupled with a Coulter S Plus II analyzer
7. Registered trade mark of Cognition Technology Corporation, Massachusetts
8. Paperback Software International, Berkeley, California
9. Available from Dr. M.L. O'Connor

There are four groups which have been working for several years on hematological diagnostic systems: a French group (Hôpital Henri Mondor, Paris; system ADH), an Italian group (University of Pavia; system NEOANEMIA), and two American groups (University of Texas, Texas; system BCDE, and University of Cincinnati, Ohio; system CONSULT-I). The other systems from Table 6.9 are either very early systems (HEME) or are developed for a small diagnostic area (microcytic anemia, thalassemia/iron deficiency, or myelodysplastic syndromes). The main characteristics of the four major systems are given in Table 6.10.

Table 6.10 *Characteristics of four major interpretative systems for anemia. n.r. means not reported.*

System	ANEMIA	ADH	BCDE	CONSULT-I
Lit.ref.	3,40	4,42,43	46	44,45
Disease entities	65	30	36	n.r.
Use of a shell	yes	yes	no	yes
Size of test set	100	180	182	84
Correct diagnosis listed as first (%)	90	65	74	64
Correct diagnosis among 1st three (%)	96	84	96	88

Table 6.10 gives the number of disease entities each system is able to diagnose. This number may have increased since the moment of publication until present. All systems except one (BCDE) use an expert system shell. The size of the test set used in the evaluation of these systems is given, which in some cases is rather small compared to the number of disease entities. The percentage of correct diagnoses in the evaluation found with the test set are given. As all systems display a list of possible diagnoses, the score for having the correct diagnosis among the first three has been included. All systems have been evaluated against human experts and performed equally or better. Hereafter, some additional characteristics of each of these systems will be discussed.

1. ANEMIA

This rule-based expert system was developed on a mainframe using the shell EXPERT. It was developed to assist the physician diagnosing anemias in an internal medicine setting. The system ANEMIA uses forward chaining (data driven reasoning) and consists of 1250 rules. It can accommodate 120 findings, among which are the results from a blood smear and from examination of the bone marrow. Its training set consisted of 200 cases of anemia. It is able to assist in the diagnosis of 65 disease entities, to provide a list of possible diagnoses and to give recommendations for management of the disease. The same group re-implemented the knowledge of ANEMIA in a system called NEOANEMIA on

an EXPLORER workstation from Texas Instruments, using the shell KEE. This was done to improve flexibility and user-friendliness.

2. ADH

The system ADH was developed on an IBM PC-XT. It was intended for adult pathology. It provides a list of possible diagnoses and it can propose follow-up tests. It can also accomodate data from a blood smear. The system was made using an expert system shell and uses the theorem of Bayes to calculate the probability of each diagnosis. Its training set consisted of 400 cases. The same group has developed an expert system for the diagnosis of myelodysplastic syndromes which additionally uses data from a bone marrow aspiration examination.

3. BCDE

This system was written in BASIC and runs on an IBM-PC compatible computer. The program accomodates data from the blood smear as well as clinical data like sex, race, and knowledge of recent chemotherapy or transfusion. The program was developed for the screening of adult patients. The system BCDE uses the results of multivariate analysis that were obtained with 8000 complete blood counts with known hematological status. The concept of 'normal blood count' was developed by multivariate analysis of 1500 blood counts from medical and nursing students. With this knowledge, a point scoring system was made indicating the probability of each diagnosis. The same group developed another program, BCDE2, that discriminates between iron deficiency and heterozygous thalassemia. When evaluated, this system outperformed both a group of human experts and two discriminant functions based on erythrocyte indices.

4. CONSULT-I

The expert system CONSULT-I was developed for an IBM PC-AT using the shell Consult Learning System. It accomodates demographical data like age, sex, race, as well as data from the blood smear, ferritin, vitamin B_{12}, and red blood cell folate. The system uses pattern recognition techniques and likelihood ratios to construct a ranked list of diagnoses. Its training set consisted of 560 cases, among which were 60 cases of mixed anemias. The system is intended to be used in screening situations. Its results are reviewed and supervised by a pathologist.

6.7 Discussion and conclusions

We have developed a prototype expert system HEMO for diagnostic support in cases of anemia and for accelerated laboratory testing. This expert system requires less diverse input than the four systems described above. In HEMO, no data from the blood smear are used, nor have factors indicating the prevalence of disease or probability of conclusions been used. Demographical data (age and sex) and laboratory data are required, but no clinical data. The system obligatory uses the Hb, RBC, MCV, and a reticulocyte count. In its 66 rules, 21 diagnostic entities have been incorporated. The system runs on a PC-

AT. Multivariate analysis is not used. Instead of creating a list of possibilities, one diagnostic interpretation is offered. The system can give advice for follow-up measurements. If there are too many unknown parameters, a list is given of the remaining diagnostic possibilities.
In a limited evaluation, the system showed a reasonable performance. Although the system is not yet finalized, many of its imperfections can easily be corrected. HEMO should be prospectively tested with a large number of policlinical cases. In this way its capability to give advice concerning additional measurements can also be tested.
An expert system for the interpretation of hematological parameters is likely to function optimally in a screening environment, but it has been shown to be of value when applied to patients in a diagnostic phase. The hemocytometric measurements provide only a rough insight in the diagnostic diversity, *e.g.* in anemia with iron and folate deficiency, normocytic as well as microcytic cases have been found. In such cases, it seems reasonable to focus on the most common signs of a particular anemia (in this case normocytosis). The choice will depend on the cost-benefit analysis of each decision. In addition, the expert system may give further recommendations, such as folate measurement in case of a low transferrin and high ferritin concentration.
In conclusion it can be said that the spin-off of this project has been most valuable: the formalization of the logic in the diagnostic process, the realization of the necessity of a reliable reticulocyte count, as well as the realization of the restrictions to the use of the MCV and especially the RDW. Furthermore, a prototype expert system has been developed, which provides a logical and standardized approach to the classification of anemias and which may eventually contribute to a better health care service to the patient.

6.8 References

1. Self KG, Conrady MM, Eichner ER
Failure to diagnose anemia in medical inpatients. Is the traditional diagnosis of anemia a dying art? Am J Med 1986;81:786-90.

2. Woo B, Jen P, Rosenthal PE, Bunn FH, Goldman L
Anemic inpatients. Correlates of house officer performance. Arch Intern Med 1981;141:1199-1202.

3. Quaglini S, Stefanelli M, Barosi G, Berzuini A
A performance evaluation of the expert system ANEMIA. Comp Biomed Res 1988;21:307-23.

4. Sigaux F, Imbert M, Priolet G, Bucquen JJ, Levy C, Sultan C
Aide à la décision en hématologie: caractéristiques et performances du programme. Press Méd 1987;16:111-4.

5. Bessman JD, McClure S, Bates J
Distinction of microcytic disorders: comparison of expert numerical-discriminant, and microcomputer analysis. Blood Cells 1989;15:533-40.

6. Hansen RM, Hanson G, Anderson T
Failure to suspect and diagnose thalassemic syndromes. Interpretation of RBC indices by the non-hematologist. Arch Intern Med 1985;145:93-4.

7. Simel DL, DeLong ER, Feussner JR, Weinberg JB, Crawford J
Erythrocyte anisocytosis. Visual inspection of blood films vs automated analysis of red blood cell distribution width. Arch Intern Med 1988;148:822-4.

8. Bessman JD, Gilmer PR, Gardner FH
Improved classification of anemias by MCV and RDW. Am J Clin Pathol 1983;80:322-6.

9. Bergin JJ
Evaluation of anemia. Getting the most out of the MCV, RDW, and other tests. Postgrad Med 1985;77:253-69.

10. Monzon CM, Beaver BD, Dillon TD
Evaluation of erythrocyte disorders with mean corpuscular volume (MCV) and red cell distribution width (RDW). Clin Pediatr 1987;26:632-8.

11. Flynn MM, Reppun TS, Bhagavan NV
Limitations of red blood cell distribution width (RDW) in evaluation of microcytosis. Am J Clin Pathol 1986;85:445-9.

12. Gardner FH
Refractory anemia in the elderly. Adv Intern Med 1987;32:155-76.

13. Simel DL
Is the RDW-MCV classification of anaemia useful? Clin Lab Haematol 1987;9:349-59.

14. Roberts GT, El Badawi SB
Red blood cell distribution width index in some hematologic diseases. Am J Clin Pathol 1985;83:222-6.

15. O'Broin SD, Kelleher BP, McCann SR, Ryder RJW, Scott JM
The value of the erythrocyte indices as a screening procedure in predicting nutritional deficiencies. Clin Lab Haematol 1990;12:247-55.

16. Uchida T
Change in red blood cell distribution width with iron deficiency. Clin Lab Haematol 1989;11:117-21.

17. Thame M, Grandison Y, Mason K, Thompson M, Higgs D, Morris J, Serjeant B, Serjeant G
The red cell distribution width in sickle cell disease - is it of clinical value? Clin Lab Haematol 1991;13:229-37.

18. Cesana BM, Maiolo AT, Gidiuli R, Damilano I, Massaro P, Polli EE
Relevance of red cell distribution width (RDW) in the differential diagnosis of microcytic anemias. Clin Lab Haematol 1991;13:141-51.

19. Marsh WL, Bishop JW, Darcy TP
Evaluation of red cell volume distribution width (RDW). Hematol Pathol 1987;1:117-23.

20. Laso FJ, Mateos F, Ramos R, Herrero F, Perez-Arellano JL, Gonzalez-Buitrago JM
Amplitud de distribución del tamaño eritrocitario en el diagnóstico diferencial de un anemia microcítica. Med Clin (Barc) 1990;94:1-4.

21. Qurtom HA, Al-Saleh QA, Lubani MM, Hassanein A, Kaddoorah N, Qurtom MA, Al-Sheikh T
The value of red cell distribution width in the diagnosis of anaemia in children. Eur J Pediatr 1989;148:745-8.

22. Miguel A, Linares M, Miguel A, Miguel-Borja JM
Red cell width distribution width analysis in differentiation between iron deficiency and thalassemia minor. Acta Haemat 1988;80:59.

23. Paterakis GS, Terzoglou G, Vasilioy E
The performance characteristics of an expert system for the "on-line" assessment of thalassemia trait and iron deficiency-Micro Hema Screen. Blood Cells 1989;15:541-61.

24. Kaye FJ, Alter BP
Red cell size distribution analysis: an evaluation of microcytic anemia in chronically ill patients. Mt Sinai J Med 1985;52:319-23.

25. van Zeben D, Bieger R, van Wermeskerken RKA, Castel A, Hermans J
Evaluation of microcytosis using serum ferritin and red blood cell distribution width. Eur J Haematol 1990;44:105-8.

26. Helleman PW, Bartels PC, van Waveren Hogervorst GD
Screening for thalassaemia using the width of the Technicon H6000/H601 erythrocyte size histograms. Scand J Clin Lab Invest 1988;48:697-704.

27. Koepke JF, Koepke JA
Reticulocytes. Clin Lab Haematol 1986;8;169-79.

28. Laharrague P, Corberand JX, Fillola G, Marcelino N
Evaluation d'un analyseur automatique de réticulocytes: le Sysmex R-1000. Ann Biol Clin 1990;48:253-8.

29. Peebles DA, Hochberg A, Clarke TD
Analysis of manual reticulocyte counting. Am J Clin Pathol 1981;76:713-7.

30. Metzger DK, Charache S
Flow cytometric reticulocyte counting with thioflavin T in a clinical hematology laboratory. Arch Pathol Lab Med 1987;111:540-4.

31. Chin-Yee I, Keeney M, Lohmann RC
Flow cytometric reticulocyte analysis using thiazole orange; clinical experience and technical limitations. Clin Lab Haematol 1991;13:177-88.

32. Green R, King R
A new red cell discriminant incorporating volume dispersion for differentiating iron deficiency anemia from thalassemia minor. Blood Cells 1989;15:481-91. Discussion ibid. 492-5.

33. Bentley SA, Ayscue LH, Watson JM, Ross DW
The clinical utility of discriminant functions for the differential diagnosis of anemias. Blood Cells 1989;15:575-82. Commentary ibid. 583-4.

34. Makris PE
Utilization of a new index to distinguish heterozygous thalassemic syndromes: comparison of its specificity to five other discriminants. Blood Cells 1989;15:497-507. Discussion ibid. 507.

35. Wiener F, de Verdier CH, Groth T
The use of knowledge-based information systems for interpreting specialized clinical chemistry analyses - experience from erythrocyte enzymes and metabolites. Scand J Clin Lab Invest 1990;50:247-59.

36. Schilling RF
Anemia of chronic disease: a misnomer. Ann Intern Med 1991;115:572-3.

37. Engle RL, Flehinger BJ
Computer-aided differential diagnosis of hematologic diseases: a Bayesian probability model. In: Izak G, Lewis SM, eds. Modern concepts in hematology. New York: Academic Press, 1972:265-72.

38. Engle RL, Flehinger BJ, Allen S, Friedman R, Lipkin M, Davis BJ, Leveridge LL
HEME: a computer aid to diagnosis of hematological disease. Bull NY Acad Med 1976;52:584-600.

39. Berzuini C, Stefanelli M
KIRON: a computer system and a methodology applicable to the study of erythroid disorders. Comp Biomed Res 1982;15:361-86.

40. Quaglini S, Stefanelli M, Barosi G, Berzuini A
ANEMIA: an expert consultation system. Comp Biomed Res 1986;19:13-27.

41. Lanzola G, Stefanelli M, Barosi G, Magnani L
NEOANEMIA: a knowledge-based system emulating diagnostic reasoning. Comp Biomed Res 1990;23: 560-82.

42. Sultan C, Priolet G, Imbert M
Logiciel d'aide à la décision. Hémogramme, myélogramme, cytologie. France, Coulter Electronics, 1987.

43. Imbert M, Priolet G, Dadi W, Sultan C
An expert system applied to the diagnosis of anemia with special reference to myelodysplastic syndromes. Blood Cells 1989;15:563-71.

44. Blomberg DJ, Guth JL, Fattu JM, Patrick EA
Evaluation of a new classification system for anemias using Consult Learning System. Comp Meth Prog Biomed 1986;22:119-25.

45. Blomberg DJ, Ladley JL, James HS, Fattu JM, Patrick EA
The use of an expert system in the clinical laboratory as an aid in the diagnosis of anemia. Am J Clin Pathol 1987;87:608-13.

46. Bates JE, Bessman JD
Evaluation of BCDE, a microcomputer program to analyze automated blood counts and differentials. Am J Clin Pathol 1987;88:314-23.

47. Goldschmidt HMJ, Akkermans AM, van Dongen CAJM
A limited experimental expert-system on anaemia diagnostics: 'RBC'. Automedica 1987;8:89.

48. Houwen B
The use of inference strategies in the differential diagnosis of microcytic anemia. Blood Cells 1989;15:509-32.

49. O'Connor ML, McKinney T
The diagnosis of microcytic anemia by a rule-based expert system using VP-expert. Arch Pathol Lab Med 1989;113:985-8.

Chapter 7

GENERAL DISCUSSION

7.1 Introduction

The subject of this thesis is to investigate how expert systems can be useful in clinical chemistry. An expert system is a computer program which can give advice in a well-defined area of expertise and is able to explain its line of reasoning. An expert system can reason with uncertainty (1).
We have developed three expert systems, all based on the empty shell 'Personal Consultant': LITHOS, CHEMPATH and HEMO. Experiments with these systems have been performed in order to test the usefulness of this offspring from Artificial Intelligence.
In clinical chemistry, expert systems can be useful in the following ways: 1. in the analytical process, 2. in the interpretation of clinical chemistry measurements, 3. in the guidance and acceleration of laboratory diagnostics, 4. in the troubleshooting of analytical instruments. An example of the first category is the expert system LITHOS, developed for the interpretation of X-ray diffractograms of urinary calculi (2, cf. ch. 2 and 3). Others have used an expert system for the validation (plausibility control) of analytical results (3). Examples of the second and third category are CHEMPATH and HEMO, expert systems which give diagnostic support in acid-base and electrolyte disorders and in anemia, respectively (4, cf. ch. 4 and 6). The aspect of acceleration of laboratory diagnostics has not been tested. Examples of the second and fourth category are given in Table 1.1 (ch.1).
Expert systems can be used within various configurations, depending on the source of the input data. We have developed a system for acid-base interpretation which functions on-line with a blood-gas analyzer (subunit of CHEMPATH, ch. 4). On request, this system generates a printed report of the acid-base interpretation, including the squared Mahalanobis distance (ch. 4). In a second configuration, the system has been adapted to retrieve a patients historical blood-gas data from the hospital information system. The successive acid-base interpretations are presented on the screen and the corresponding squared Mahalanobis distances are graphically displayed as a time series. In a third configuration, the data are retrieved from a laboratory database. We have developed a Turbo Pascal program functioning on a local area network, which fills a laboratory database with blood-gas or blood-count results. This database was intended as a source of input data for an expert system. For such a function, data from one patient obtained with several analytical instruments should be automatically collected *via* a common

characteristic, *e.g.* the patient identification number. Because this number was inaccessible to the computer for privacy reasons, its use was precluded.

7.2 Diagnostic support

The automatization of clinical chemistry laboratories has led to an increase in capacity. The amount of data that is produced nowadays is too large to be interpreted by the clinical biochemist, in the way it was traditionally. Modern techniques of data processing offer the possibility of an automated interpretation of the analytical data, in which expert systems can play an important part.
A current misconception is that this interpretation is an attempt to make a diagnosis. A diagnosis is usually based on information from various sources, *e.g.* clinical, chemical, radiological and microbiological information. Only in exceptional cases is the information provided by the clinical chemistry laboratory sufficient to make a diagnosis. Therefore, the use of expert systems in clinical chemistry is certainly not meant to provide diagnoses.
There are, however, forms of interpretation which can be generated from only laboratory measurements. The recognition of a microcytic hypochromic anemia is not a diagnosis, but a conclusion which can be validly drawn from laboratory measurements. Providing information about the association of this observation with various diseases is still within the professional domain of clinical biochemistry. The challenge to clinical chemistry is to manoeuvre in the narrow strait between numbers and diagnoses to extract the maximum possible amount of information from the primary data.
The enhanced information must not be forced upon the attending physician. However, the human mind is not perfect at remembering, and especially in overloaded consulting hours, details are at risk of being overlooked. Especially younger, less experienced physicians, which are usually in charge at night or in weekends when expert opinion is not readily available, will be well served with an interpretative report. In the opinion of the author, the ideal situation should bear resemblance to the traditional situation. The physician who wants to consult the clinical biochemist dials his number (pushes a button on his terminal) and is presented with diagnostic support, based on the results from clinical chemistry measurements. In case of further questions, a personal contact should be preferred over an automatized explanation.

7.3 Reasoning with uncertainty

Expert system shells usually offer the possibility of reasoning with uncertainty. This is done by the use of certainty factors, representing a degree of confirmation of a certain hypothesis, based on the combination of an *a priori* probability and new evidence. Shortliffe and Buchanan introduced the certainty factor as the difference between the

increased belief in a hypothesis and the increased disbelief in the same hypothesis, both as a result of the new evidence (5).

At least four kinds of uncertainty can be found in the diagnostic process:

1. Every measurement in clinical chemistry is subject to a small amount of *analytical* variation, which is known from quality control measurements (the preanalytical variation is unknown, but small). The analytical variation is always small compared to the variabilities mentioned hereafter. We have proposed a calculation of certainty factors for this type of uncertainty (4).
2. Every measurement has an amount of *physiological* variation within a healthy subject (circadian, seasonal, or irregular fluctuations). There is also an inter-individual variation between healthy subjects. These types of variation are usually not considered separately, and are included in the traditional reference range.
3. Every parameter displays a *pathological* variation which is a resultant of the type of disease, its stage and the characteristics of the individual patient.
4. Every diagnostic conclusion is subject to an amount of *diagnostic* uncertainty, caused by incomplete evidence and by the semi-quantitative nature of the evidence on which the diagnosis is based, and by the fact that many diseases have no pathognomonic characteristics.

In well-defined situations (1 and 2), the use of the frequency ratio is recommended to calculate *e.g.* the probability that a given result is within its reference limits. The uncertainties mentioned in 3 and 4 are to a large extent unknown, or can only be inaccurately estimated. It will be clear that the incorporation of these uncertainties in our expert systems would be a subjective and highly complicated undertaking. Therefore we have refrained from the use of certainty factors. In LITHOS, where the input data were very precise, the only uncertainty was a minor between-run shift of peak positions.

7.4 Techniques

Examples of the techniques, which can be used to produce enhanced information are the use of discriminant functions and pattern recognition. Pattern recognition techniques are usually based on Bayes' theorem. These techniques or the results thereof can be incorporated into expert systems to assist in the reasoning process. The barrier to using these techniques is not their unavailability, but the lack of familiarity with them. The advantage in using these techniques is that clinical chemistry measurements are not considered as isolated results, but taking their interdependency into account.

We have developed a multivariate model for acid-base measurements (6,7, ch. 5). From this model, the Mahalanobis distance can be calculated, which quantitatively characterizes the acid-base condition of the patient relative to the mean of the hospital population. The monitoring of this distance offers the possibility to determine the relative stability of the acid-base disorder. In addition, this approach ensures a standardized interpretation. Its

performance in detecting acid-base disturbances was not grossly different from that of human experts (ch. 4). This model is currently being routinely used for acid-base monitoring in another hospital (Onze Lieve Vrouwe Gasthuis, Amsterdam).

7.5 Programming tools

Although the use of an expert system shell facilitates programming and reasoning with uncertainty, it is not obligatory. Many of the published systems have been written in conventional programming languages. In some cases systems have been developed using a shell, and were later translated into a conventional language to limit their cost and to speed up their performance (8). We have compared an expert system in LISP with a program in Turbo Pascal for the interpretation of X-ray diffractograms of urinary calculi, and we found that in this analytical setting both programs produce equal results in equal periods of time (ch. 3).
The use of an expert system shell results in an expert system that is able to explain its line of reasoning. The advantage of this property is usually largely overestimated and it is of no use in a fully automated situation, because there is no need for an explanation. The main advantage of a shell is that it facilitates the development phase of an expert system. Decision rules can be easily added, removed, or rearranged. A trace option is available, which shows whether rules are applied or rejected, and in what sequence. Such a trace can be stored on disk and consultations can be rerun with a minor change in input data, to study the effect of such a change.

7.6 Knowledge elicitation

The development of expert systems necessitates its developers to think thoroughly about the knowledge domain. It is precisely this process which makes it painfully clear how little hard-core knowledge there really is available. The knowledge needed to create such systems can be obtained in several ways.
1. Knowledge can be extracted by interviewing experts on the subject, which is usually a very tedious procedure. The experts must have a consensus opinion, and the reasoning process has to be explained in great detail to the knowledge egineer (usually a layman) who is to implement the expertise. Clinical biochemists are in an advantageous situation because they are experts in their professional domain and they are familiar with the use of computers for data communication and data processing. In this way they are in an ideal position to develop expert systems themselves.
2. Knowledge can be obtained from large collections of data by numerical analysis. If demographic or clinical data are additionally available, the interpretative possibilities are increased. Unfortunately, the retrieval of clinical data usually means tedious work-up of patient records.

3. Knowledge can be obtained from the literature. This must be done with caution in order not to arrive at faulty conclusions (ch. 6, section 6.2.1). The author has been somewhat shocked at the amount of confusion in the literature over the interpretation of only one parameter (RDW) in three of the most common causes of anemia (ch. 6, section 6.2.1.1). A clear picture in these cases can only be obtained by an overview of the literature, which accentuates the importance of such transversal studies.
The knowledge in LITHOS and HEMO was largely obtained from interviewing experts (and from experiments in the case of LITHOS), whereas the knowledge in CHEMPATH was obtained from the literature and from data analysis.

7.7 Experiences with expert systems

The development phase of expert systems can have useful spin-off. In the design of CHEMPATH we found a large number of acid-base measurements which are unclassifiable by the traditional univariate classification. This led to the development of a multivariate model based on a patient population (ch. 4,5), in order to detect acid-base disturbances. The development of HEMO made us realize the relativity of the diagnostic value of the red cell indices and the importance of an accurate and precise reticulocyte count (ch. 6, section 6.2.1). The use of LITHOS and CALCULI demonstrated that for the interpretation of X-ray diffractograms a conventional program (in Turbo Pascal) and an expert system perform equally well and comparable to the traditional method (ch. 3).
The first version of the empty shell Personal Consultant was altered to enable manipulation of the system by LISP programs, in order to avoid the use of programs in pseudo-language with an English explanation. These would have occupied an undesirable proportion of the internal memory and thereby resulted in a slower expert system. The incorporation of LISP programs in the rules mutilated the systems ability to explain its reasoning, because no translation of these programs into English could be given. However, the use of LISP provided us with all the facilities of a high level language, which proved to be indispensible in the construction of our expert systems. We have written programs which, during a consultation, alter the premise of certain rules, programs which automatically incorporate a debugging facility in another program, and even programs which on request construct the basic skeleton of other programs. These programs were incorporated in LITHOS, in order to extend the possibilities of the first version of the empty shell.
However, we found our expert systems in LISP to be rather slow. By writing the LISP code in LITHOS more efficiently, we were able to double the speed of its execution. Nowadays there is a trend towards the use of faster shells in PROLOG or in C. In the development of LITHOS, the 640 kb internal memory of the Olivetti M24 personal computer became a limitation when the system contained approximately 100 decision rules. The frequent 'garbage collections' of 30 seconds during which the system was locked while 'cleaning

up' its memory were an annoying feature. In a newer version of the empty shell, the memory limitation is less severe because only the code which is directly needed is loaded into the internal memory. Also, the periods of garbage collection are much shorter and much less a nuisance. In view of the above arguments, the use of expert system shells in LISP is not recommended.
We had an unpleasant experience when learning that the manufacturer of the Personal Consultant empty shell did no longer support this product in Europe because lack of interest from the market.

7.8 Advantages and disadvantages of expert systems

Expert systems offer the continuous availability of expertise. They can save time for the human expert. They offer a standardized and logical approach to their problem area. They can be used in analytical situations (ch. 2), in plausibility control, in instrument troubleshooting, in diagnostic support (ch. 4,6), and in accelerated laboratory testing.
A disadvantage is that in expert systems only incorporated variates will be considered, whereas the human mind tends to take additional information into account and to weigh the various factors according to the situation. This is a principal, yet not a decisive argument, because any interpretative system can be expanded to include additional information. An equally indecisive argument is the fact that expert systems do not operate flawless, because this holds equally for human experts, as reported in the literature.

7.9 Practical considerations

The introduction of an expert system in its operational environment should be done gradually. By analogy to the introduction of a new analytical instrument, a period of time in which the old and the new system function parallel, is advantageous. In this way, confidence in its performance will be created in all persons involved with its function. An expert system should be introduced only after thorough testing by its developers, in order to prevent premature scepticism. Its functioning should be continuously be monitored and evaluated as it should be, but seldom is, with human experts.
The formalized knowledge in an expert system will be commercially attractive. A discussion on the caution that has to be applied in the transfer of expert systems, has been given in chapter 1. In the new environment, the evaluation phase should be repeated, and maintenance and legal aspects should be strictly regulated. The incorporation of expert knowledge in analytical instruments for diagnostic purposes is a development in which clinical biochemists should participate.

7.10 Neural networks

A further development in artificial intelligence is that of neural networks. These are programs for pattern recognition, which have the advantage of being more flexible than expert systems, since no decision tree is needed. Disadvantages are their inability to give advice, to explain their reasoning, to give selectively weight to certain conclusions, as well as their sensitivity to a correct choice of the training set and to dimensioning. Comparisons have been made between the use of neural networks and of multivariate analysis in removal of effects of interfering compounds in colorimetry (9), in the diagnosis of breast carcinoma from laboratory data (10), and in the grading of breast carcinoma (11), in which both methods functioned equally well. In our opinion, neural networks are used preferably in situations where logical reasoning is difficult to apply, where no interaction is needed, and where there are subtle nuances to be differentiated. Recent research is directed towards an integrated use of expert systems and neural networks.

7.11 Conclusions

From the literature examples in this thesis it is clear that much research has already been done on the use of expert systems in clinical chemistry. It is also clear that research in the Netherlands in this area is lagging behind. We have shown that it is possible and desirable to develop expert systems that offer analytical or diagnostic assistance. The techniques to do this are unfamiliar in clinical chemistry in the Netherlands. A course of multivariate statistics and expert systems in the education of clinical biochemists is recommended to enable a widespread application of these techniques.
It is hoped that this thesis has shown the need for interpretative reporting. The data, statistical techniques, and programs that incorporate such techniques are available. In short, all conditions are fulfilled to create a better information service, which ultimately will be advantageous to the patient.

7.12 References

1. Leijnse B
Het expertsysteem. Een hybride van een orakel en een geheugensteuntje. COBO-bulletin 1989;22:6-16.

2. Wulkan RW, Zwang L, Liem TL, Blijenberg BG, Leijnse B
Renal stone analysis: LITHOS, an expert system for evaluation of X-ray diffractograms of urinary calculi. J Clin Chem Clin Biochem 1987;25:719-22.

3. Valdiguié PM, Rogari E, Philippe H
VALAB: Expert system for validation of biochemical data. Clin Chem 1992;38:83-7.

4. Wulkan RW, Leijnse B
Experience with expert systems in clinical chemistry. In: Kerkhof PLM, van Dieijen-Visser MP, eds. Laboratory data and patient care. New York: Plenum Press, 1988:117-24.

5. Shortliffe EH, Buchanan BG
A model of inexact reasoning in medicine. Math Biosci 1975;23:351-79.

6. Gelsema ES, Leijnse B, Wulkan RW
Detection of aberrant observations in a background of an unknown multidimensional Gaussian distribution. Meth Inform Med 1990;29:236-42.

7. Gelsema ES, Leijnse B, Wulkan RW
A multi-dimensional analysis of three chemical quantities in the blood. Med Inform 1991;16:43-54.

8. Drouen A, Dolan JW, Snyder LR, Poile A, Schoenmakers PJ
Software for chromatographic development. LC-GC International 1992;5:28-37.

9. Burtner K, Frye S
Use of serum blank information to quantify chromogenic interferents and correct sensitive analyses. Clin Chem 1990;36:1584-6.

10. Astion ML, Wilding P
Application of neural networks to the interpretation of laboratory data in cancer diagnosis. Clin Chem 1992;38:34-8.

11. Dawson AE, Austin RE, Weinberg DS
Nuclear grading of breast carcinoma by image analysis. Classification by multivariate and network analysis. Am J Clin Pathol 1991;95 (suppl 1):29-37.

SUMMARY

Aim of this thesis

The aim of this thesis is to investigate where and how expert systems can be used for information enhancement in clinical chemistry. It has been written for a public of clinical biochemists with the hope to convert the existing scepticism about expert systems into enthousiasm.
The content of this thesis is multifacetted by nature. In this aspect it differs from the usual type of thesis. The reader should realize, that precisely this multifacetted character is the leading thread of the thesis, and *not* an in-depth discussion of the diverse subjects.

Chapter 1: Introduction

In this chapter, the general aspects of expert systems and multivariate analysis are discussed. It is explained that the growing capacity of the clinical chemical laboratory enforces the need for automatic data interpretation. The traditional interpretation by comparison with a reference range is discussed, as well as its drawbacks when applicated to multi-analytical measurements. In this area, multivariate analysis can offer a number of advantages over the traditional approach. The need for information enhancement in clinical chemistry is illustrated with examples from the literature.
A short review of the history of expert systems is given, and the languages, shells, structure and function of expert systems are discussed. Also, a discussion is given of some problems that can be encountered in developing expert systems.
The juridical aspects of making and using expert systems are briefly explained. An extensive list of applications of expert systems in clinical chemistry is presented. Finally, a short discussion is given of future prospects in expert systems, and the scope of this thesis is given.

Chapter 2: LITHOS

In this chapter, an expert system is described which now is succesfully operative in the clinical chemistry laboratory. The expert system LITHOS interprets X-ray diffractograms of urinary calculi and reports the qualitative and quantitative compositions. The system

was developed at a time when analytical capacity had to be increased and human expertise was scarce. A description is given of the program and the flow of the inference process. For the interested reader, a detailed example of a complex consultation is available on request from the author of this thesis.

Chapter 3: LITHOS and CALCULI

This chapter is written in the form of a publication, because it is yet to be submitted. Two computer programs for the interpretation of X-ray diffractograms are compared. The first (LITHOS) is an expert system that has been described in chapter 2, and the other (CALCULI) has been developed at the Technical University Delft using a conventional programming language. The performance of both systems are compared to the results with the conventional Debije-Scherrer-Hull technique. The reproducibility of the analytical procedure including the interpretation by both systems has been investigated for three frequently occurring components. By further experiments, the main source of variation was found to be the analytical instrument, and not the sample preparation or the programs interpretation. A description is given how one weaker aspect in the interpretation of CALCULI could be improved. The problem of a new, unknown component is illustrated by the performance of LITHOS with urinary calculi from Indonesia. After a comparison of consultation times, the chapter closes with some final remarks, describing the incorporation of both systems in the analytical procedure. When a consensus is reached by both programs, the result is reported to the attending physician. This is the majority of cases, which thus can be handled without human expertise. When there is a difference of opinion, the human interpretation of an X-ray film made with the Debije-Scherrer-Hull camera has the final word.

Chapter 4: CHEMPATH

This chapter begins with an ample discussion of the nomenclature for acid-base disorders. When this aspect has not been clearly defined, misunderstanding may arise easily. Several illustrative cases of acid-base disturbances are presented. A system of nomenclature is proposed, and tested with well-defined literature cases. Knowledge of the anion gap would enhance the interpretative power of an automated system. Further handicaps are missing information (artificial ventilation, medication) and organizational factors.
Examples taken from the literature are presented illustrating the need for intrepretative systems. An overview is given of existing interpretative systems for acid-base disorders and their characteristics. The expert system CHEMPATH for acid-base and electrolyte disturbances is described.

Chapter 5: Multivariate analysis

This chapter describes the use of multivariate analysis to calculate reference ranges from patient data. These reference ranges can be used to signal early pathology and to detect laboratory error. A discussion is given of the advantages and disadvantages of this approach. The theory and calculation of these reference ranges is discussed. Results are presented of the calculation of patient-based reference ranges for acid-base parameters for different age groups. A discussion of these results in relation to the concept of setpoint is given. The use of the multivariate reference range is compared with the judgement of two clinicians. The chapter ends with a discussion of the multivariate changes of acid-base parameters in time.

Chapter 6: HEMO

This chapter describes the development of an expert system for the interpretation of anemia. First, the diagnostic value of the MCV, RDW and reticulocyte count is discussed. An overview of RDW ranges in three pathological conditions is given, showing large overlap. A short discussion of discriminant functions is given.
The development of HEMO is described, beginning with an analysis of the pattern of requests with policlinical patients. The structure of HEMO is described next. An overview is given of interpretative systems for anemia, with a focus on the work of four international groups. The need for interpretative reporting is illustrated with examples from the literature. Finally, the system is evaluated with a large set of patient data.

Chapter 7: General discussion

This last chapter presents a discussion of the application of expert systems in clinical chemistry. The functions that such systems can have in the clinical chemistry laboratory are mentioned, with examples. The character of the diagnostic support, that expert systems for clinical chemistry should give, is described. The sources of uncertainty in the diagnostic process are given, and it is explained why no use has been made of certainty factors. The techniques to produce enhanced information are mentioned and the use of an expert system shell is discussed. Several possible sources of knowledge for expert systems are mentioned. Experiences with expert systems are reported, as well as a discussion of advantages and disadvantages. After some practical considerations, a short section about neural networks is presented. The chapter closes with some conclusions and recommendations.

SAMENVATTING

Doel van dit proefschrift

Het doel van dit proefschrift is te onderzoeken waar en hoe expertsystemen kunnen worden gebruikt voor een verbeterde informatievoorziening vanuit de klinische chemie. Het is geschreven voor een publiek van klinisch chemici in de hoop de bestaande scepsis ten aanzien van expertsystemen in enthousiasme om te zetten.
De inhoud van dit proefschrift is van nature veelzijdig. In dit opzicht verschilt het van het normale type proefschrift. De lezer dient zich te realiseren, dat juist deze veelzijdigheid de rode draad door dit proefschrift vormt, en *niet* een diepgaande discussie van de verscheidene onderwerpen.

Hoofdstuk 1: Inleiding

In dit hoofdstuk worden de algemene aspecten van expertsystemen en multivariate analyse besproken. Er wordt uitgelegd dat de groeiende capaciteit van het klinisch-chemisch laboratorium de noodzaak van automatische interpretatie van de gegevens doet toenemen. De traditionele interpretatie door vergelijking met een referentiegebied wordt besproken, evenals de nadelen hiervan bij toepassing op verschillende analyses. Hier kan de multivariate analyse een aantal voordelen bieden ten opzichte van de traditionele aanpak. De noodzaak van een verbeterde informatievoorziening vanuit de klinische chemie wordt geïllustreerd met voorbeelden uit de literatuur.
Er wordt een kort overzicht van de geschiedenis van expertsystemen gegeven en de talen, shells, structuur en functies van expertsystemen worden besproken. Ook wordt een aantal problemen besproken die zich kunnen voordoen bij het ontwikkelen van expertsystemen. De wettelijke aspecten van het maken en het gebruiken van expertsystemen worden kort van commentaar voorzien. In een uitgebreide lijst worden toepassingen van expertsystemen in de klinische chemie vermeld. Tenslotte wordt een korte discussie gegeven van de toekomstige ontwikkelingen op het gebied van expertsystemen, alsmede een afbakening van het onderwerp van dit proefschrift.

Hoofdstuk 2: LITHOS

In dit hoofdstuk wordt een expertsysteem beschreven dat succesvol wordt gebruikt in het klinisch-chemisch laboratorium. Het expertsysteem LITHOS interpreteert röntgen-diffractogrammen van urinewegstenen en rapporteert de kwalitatieve en kwantitatieve samenstellingen. Het systeem is ontwikkeld toen de analytische capaciteit vergroot moest worden en de menselijke expertise schaars was. Er wordt een beschrijving gegeven van het programma en van de redeneerwijze daarvan. Voor de geïnteresseerde lezer is er een gedetailleerd voorbeeld van een ingewikkelde consultatie op verzoek verkrijgbaar bij de auteur van dit proefschrift.

Hoofdstuk 3: LITHOS en CALCULI

Dit hoofdstuk is geschreven in de vorm van een publicatie; als zodanig zal deze nog aangeboden worden. Er worden twee computerprogramma's voor de interpretatie van röntgen-diffractogrammen vergeleken. Het eerste (LITHOS) is een expertsysteem dat in hoofdstuk 2 is beschreven, en het tweede (CALCULI) is ontwikkeld in een conventionele programmeertaal aan de Technische Universiteit Delft. De verrichtingen van beide systemen worden vergeleken met de resultaten van de techniek volgens Debije, Scherrer en Hull. De reproduceerbaarheid van de analytische procedure, met inbegrip van de interpretatie van beide systemen, is onderzocht voor drie frequent voorkomende componenten. Uit verdere experimenten blijkt dat de belangrijkste bron van variatie het analytisch instrument is, en niet de monstervoorbereiding of de interpretatie van het programma. Er wordt beschreven hoe een zwakker aspect van de interpretatie van CALCULI kon worden verbeterd. Het probleem van een nieuwe, onbekende component wordt geïllustreerd door het gedrag van LITHOS met urinewegstenen uit Indonesië. Na een vergelijking van consultatietijden sluit het hoofdstuk af met enkele opmerkingen omtrent de inbedding van beide systemen in de analytische procedure. Als de systemen een consensus bereiken wordt het resultaat gerapporteerd aan de behandelende arts. Dit gebeurt in het grootste aantal van de gevallen, waardoor deze zonder menselijke expertise kunnen worden afgehandeld. Als de resultaten verschillen, geeft de interpretatie door een menselijke expert van een röntgenopname met de Debije-Scherrer-Hull-camera de doorslag.

Hoofdstuk 4: CHEMPATH

Dit hoofdstuk begint met een uitgebreide discussie van de nomenclatuur voor zuur-base-afwijkingen. Indien dit aspect niet duidelijk is gedefinieerd kunnen er gemakkelijk misverstanden ontstaan. Verschillende illustratieve gevallen van zuur-base-afwijkingen worden gepresenteerd. Er wordt een systeem voor nomenclatuur voorgesteld en getest met

goed gedefinieerde *casus* uit de literatuur. Kennis van de grootte van de anion gap zou de interpretatieve kracht van een dergelijk geautomatiseerd systeem kunnen vergroten. Verdere handicaps worden gevormd door ontbrekende informatie (kunstmatige beademing, medicatie) en door organisatorische factoren.
Er worden voorbeelden uit de literatuur gegeven die de noodzaak van interpretatieve systemen illustreren. Ook wordt een overzicht gegeven van de bestaande systemen voor zuur-base-afwijkingen en hun karakteristieken.

Hoofdstuk 5: Multivariate analyse

Dit hoofdstuk beschrijft het gebruik van multivariate analyse om referentiegebieden te berekenen uit patiëntengegevens. Deze referentiegebieden kunnen worden gebruikt om vroege pathologie te signaleren en om fouten van het laboratorium te detecteren. De voor- en nadelen van deze benadering worden besproken, alsmede de theorie en de berekening van genoemde referentiegebieden. De resultaten worden gepresenteerd van een berekening van referentiegebieden voor zuur-base-parameters, op basis van patiëntengegevens voor verschillende leeftijdsgroepen. Deze resultaten worden besproken in samenhang met het begrip 'setpoint'. Het gebruik van multivariate referentiegebieden wordt vergeleken met het oordeel van twee clinici. Het hoofdstuk eindigt met een bespreking van multivariate veranderingen van zuur-base-parameters in de tijd.

Hoofdstuk 6: HEMO

Dit hoofdstuk beschrijft de ontwikkeling van een expertsysteem voor de interpretatie van anemie. De diagnostische waarde van de MCV, de RDW en het aantal reticulocyten wordt besproken. Een overzicht wordt gegeven van RDW-waarden in drie pathologische toestanden, waaruit een grote overlap blijkt. Er wordt een korte bespreking van discriminant-functies gegeven.
De ontwikkeling van HEMO wordt beschreven, te beginnen met een analyse van de aanvraagpatronen bij poliklinische patienten. Vervolgens wordt de structuur van HEMO aangegeven. Een overzicht van interpretatieve systemen voor anemie wordt gegeven, met extra aandacht voor het werk van vier internationale groepen. De noodzaak van interpretatieve rapportage wordt geïllustreerd met voorbeelden uit de literatuur. Tenslotte wordt het systeem geëvalueerd met een groot aantal patiëntengegevens.

Hoofdstuk 7: Algemene discussie

Dit laatste hoofdstuk biedt een algemene beschouwing over de toepassing van expertsystemen in de klinische chemie. De functies die dergelijke systemen in het klinisch-chemisch laboratorium kunnen vervullen worden genoemd, met voorbeelden. Het

karakter van de diagnostische ondersteuning, die expertsystemen in de klinische chemie zouden moeten geven, wordt beschreven.
De bronnen van onzekerheid in het diagnostisch proces worden genoemd en er wordt uitgelegd waarom er geen gebruik van 'certainty factors' is gemaakt. De technieken om verbeterde informatie te produceren worden genoemd en het gebruik van expertsysteem-shells wordt besproken. Verschillende bronnen van kennis voor expertsystemen worden genoemd. De eigen ervaringen met expertsystemen worden vermeld, tezamen met een opsomming van voor- en nadelen. Na enkele praktische opmerkingen volgt een kort onderdeel over neurale netwerken. Het hoofdstuk sluit af met conclusies en aanbevelingen.

DANKWOORD/ACKNOWLEDGEMENT

In de eerste plaats wil ik mijn promotor Prof. Dr B. Leijnse bedanken voor de vele stimulerende discussies en zijn nimmer aflatend enthousiasme voor mijn promotie-onderwerp. Ik meen veel van u geleerd te hebben. Mijn tweede promotor Prof. Dr E.S. Gelsema dank ik voor zijn evenwichtige en zorvuldige begeleiding bij het schrijven van dit proefschrift en voor het opstellen van het multivariate model. Dr J. Lindemans dank ik voor zijn kritische kijk op het onderwerp, waardoor steeds weer verbeteringen mogelijk bleken, voor het leveren van de expertkennis voor HEMO, en voor de prettige samenwerking in het Sophia Kinderziekenhuis, waar hij mij in de gelegenheid stelde om naast mijn werk als klinisch chemicus dit proefschrift af te ronden. Prof. Dr P.F. de Vries Robbé, arts ben ik erkentelijk voor zijn bereidheid om deel uit te maken van de zogenaamde 'kleine commissie'.

Dan gaat mijn uitdrukkelijke dank naar de heer L. Zwang, ing. voor de vele jaren van uiterst plezierige en creatieve samenwerking, die onder andere uitmondde in de ontwikkeling van LITHOS, zover mij bekend het eerste expertsysteem in Europa dat in de klinische chemie werd toegepast.

Met respect wil ik hier ook de heer T.L. Liem gedenken, die niet alleen een prettige en zorgvuldige collega was, maar tevens een belangrijk deel van de expertkennis voor LITHOS leverde.

Drs B. van den Berg, arts dank ik voor zijn positieve opstelling en bijdrage aan het evalueren van het multivariate model voor zuur-base gegevens.

Ir A.J. van den Berg en de heer E.J. Sonneveld dank ik voor het ter beschikking stellen en aanpassen van het programma CALCULI.

Mevr. drs E.J. Harthoorn-Lasthuizen, arts dank ik voor het ter beschikking stellen van de gegevens van haar onderzoek naar anemiëen. Dr C. Vreugdenhil, arts dank ik voor het ter beschikking stellen van de gegevens van zijn onderzoek naar patiënten met rheumatoïde arthritis.

Ir drs J.J. Hermans, arts dank ik voor zijn bijdrage, in het prille stadium, aan de ontwikkeling van HEMO. Ook gaat mijn dank naar de stagiaires van het HLO-West Brabant, die elk hun steentje bijgedragen hebben aan een deel van dit proefschrift: de heren E.E.L. Keim, ing., W.C.A.M. de Volder, ing., J.A.M. Huybregts, ing., C.W.M. Heshof, ing., en F.J.M. Vermeulen, ing.

Verder rest mij nog al diegenen te bedanken die een directe of indirecte bijdrage aan het ontstaan van dit proefschrift hebben gehad. Hiervan verdienen speciale vermelding de heren A. van der Tas, R. Brouwer, M. van Vliet en mevrouw M.I.E. Huijskes-Heins.

CURRICULUM VITAE

R.W. Wulkan werd op 21 juli 1954 geboren te Pretoria (Zuid-Afrika). Kort daarna keerden zijn ouders terug naar Nederland. Na het doorlopen van de H.B.S.-B volgde een studie in de scheikundige technologie aan de Technische Universiteit Delft. Tezelfdertijd studeerde hij viool aan het Koninklijk Conservatorium in Den Haag en het Sweelinck Conservatorium in Amsterdam. Na het behalen van het doctoraal examen werd een aanstelling als wetenschappelijk assistent aan de afdeling Chemische Pathologie van de Erasmus Universiteit Rotterdam verkregen. Na deze periode volgde een opleiding tot klinisch chemicus in het Academisch Ziekenhuis Rotterdam, die werd afgesloten met inschrijving in het Register als erkend klinisch chemicus. Gedurende een periode van twee jaar werd ervaring in deze functie opgedaan in het Sophia Kinderziekenhuis te Rotterdam.

Het hier beschreven onderzoek werd verricht op de afdeling Chemische Pathologie van de Erasmus Universiteit en in het Centraal Klinisch Chemisch Laboratorium van het Academisch Ziekenhuis Rotterdam-Dijkzigt. Het onderzoek over de multivariate analyse werd verricht in samenwerking met de afdeling Medische Informatica van de Erasmus Universiteit Rotterdam.