CLIN. CHEM. 38/1, 83–87 (1992)

VALAB: Expert System for Validation of Biochemical Data

Pierre M. Valdiguié,[1] Eric Rogari,[1,2] and Hervé Philippe[3]

In large laboratories that use "high-throughput" equipment, it is now possible to use artificial intelligence techniques to aid decision making and validation of data. This paper describes an artificial intelligence project, VALAB, that has been carried out in our laboratory. VALAB, an expert system that permits real-time validation of data, is designed to be equivalent to validation by the laboratory director. The decision produced by the expert system is based on several factors, including correlation between repeated laboratory results, physiological association between different variables, the hospital department from which the test was ordered, and the patient's age and sex. In 200 abnormal chemistry profiles randomly selected, VALAB's ability to detect abnormal cases (i.e., sensitivity = 0.75) was exceeded by only one of seven laboratory experts. However, all seven experts outperformed VALAB's measured specificity of 0.63. The VALAB system incorporates >4000 rules. Operational since November 1988, it has validated >50 000 medical patients' reports in real time.

The need for high degrees of quality control in laboratories with large workloads is now well established. Moreover, the implementation of artificial intelligence techniques (*1*, *2*) within this environment is growing rapidly, particularly where such systems provide the clinical pathologist or laboratory supervisor with help in validating the final laboratory report.

An expert system, designated VALAB (*V*alidation *A*ssistee in the *LAB*oratory), is a new concept in computer-assisted validation. The system was conceived and developed in a large hospital laboratory that processes more than 1000 patients' specimens per day. The system is designed to validate biochemical profile results based on the amount of change in repeated tests, comparison of physiologically related analytes, hospital location, and patient's age and gender. The expert system was first designed for an electrolyte profile (*3*) but has been expanded to handle 22 tests commonly run in the clinical chemistry laboratory. The system is used in routine operation, is integrated into laboratory hardware, and provides an autonomous, real-time assessment of data that is the equivalent of validation by a pathologist.

[1] Laboratoire de Biochimie, Centre Hospitalier Universitaire de Rangueil, 1 avenue J. Poulhes, 31054 Toulouse Cedex, France.
[2] Société EREMS, Z.I. de Flourens, 31130 Balma, France.
[3] Laboratoire d'Automatique et d'Analyse des Systèmes, CNRS, 7 avenue du Colonel Roche, 31400 Toulouse, France.

Presented in part as a poster at XIV Int. Congress of Clin.Chem., San Francisco, CA, July 23, 1990 (Clin Chem 1990;36:1015, abstract 295).
Received December 17, 1990; accepted October 23, 1991.

Materials and Methods

Equipment

The VALAB system operates on a microcomputer and is connected with a mainframe computer (LM2 from Technicon), which treats it as an intelligent work station.

The microcomputer is an IBM-compatible PC-AT (Compak, Microdis, 31700 Blagnac, France) containing an Intel 80286 or 80386 processor, 640 kilobytes of RAM, a 40-megabyte hard disk, and Hercules or VGA graphics. The software runs under MS-DOS. It uses an expert system generator (inference engine) KHEOPS from the Laboratoire d'Automatique et d'Analyse des Systèmes, an institute of the Centre National de la Recherche Scientifique in France. In KHEOPS, forward chaining is used as the reasoning process that is applied to knowledge represented in the form of production rules or propositional logic (*4*). Groups of rules are compiled, resulting in the construction of a decision network that can be more rapidly processed than the rules in their original form. During the compilations process, the internal consistency of sets of rules is evaluated.

Methods

Contents of the knowledge base

1. Knowledge acquisition. The acquisition of knowledge by such a system must be carried out carefully and requires the services of an individual skilled in information science, because the syntax for rule writing is contained in a specialized computer language similar to LISP. Interviews, discussion with the experienced medical staff of the laboratory, and adaptation of data from the literature were used for knowledge acquisition before its integration within the software.

2. Analytical steps. Like many laboratory supervisors, VALAB uses the following information to help decide whether to validate laboratory data (*3*): (*a*) Comparison between the present and preceding results (delta check) (*5*) with calculations of a stability coefficient (the stability coefficient for any analyte is the ratio of the current value to the preceding value); (*b*) correlation between data from analytes that are physiologically linked (e.g., urea–creatinine or sodium–chloride); (*c*) ionic balance (verification of the electroneutrality between ions); (*d*) the patient's location within the hospital, because the disease is usually related to a medical specialty (e.g., renal failure in the nephrology department); (*e*) the patient's age and sex (e.g., for alkaline phosphatase or uric acid).

3. Qualitative reasoning. Acceptability thresholds are defined dynamically (for each patient) as various trends for that patient are noted. This is inspired from

the "qualitative reasoning" approach and the second generation of expert systems. Thus, the acceptable amount of abnormality in an analyte value depends on the weight of the evidence for a particular diagnosis. For example, a serum potassium value of 6.5 mmol/L will be accepted for a patient who has other evidence of chronic renal failure (e.g., other analyte values consistent with renal failure, or the specimen coming from a nephrology ward).

Construction of the knowledge base

1. Knowledge representation. Knowledge is represented by about 4500 production rules, each of which is expressed in conditional (if–then) form. Each rule contains a premise, a conclusion, and an action. For example, rule 322 is written as follows:

If there is a request for serum potassium,
And there is no preceding result,
And the potassium is abnormal (>5.8 mmol/L),
And the plausibility criteria (high creatinine, high urea, high uric acid, low calcium, specimen coming from nephrology ward or intensive-care unit) are present,
Then this abnormal potassium value is accepted.

2. Structure of the knowledge base. Structured to allow for optimal performance and easy maintenance, the knowledge base is organized as follows. The global knowledge base (~4500 rules) is divided into 69 smaller units, known as rule groups. Each rule group contains between 50 and 200 rules, which are related to a similar topic. These rule groups are placed at various locations on a decision tree. Conclusions from a particular part of the tree can be transmitted for use at other locations in the trees.

A test for a particular analyte has several rule groups associated with it. For example, one group checks the present value of the analyte against the previous value. A second rule group makes a conclusion regarding the hospital location of the patient. A third rule group interprets the analyte value with respect to other laboratory information. A final rule group incorporates information from the other rule groups and makes a decision regarding the acceptability of the test result.

3. Compilation of the rule bases. This is a "pretreatment" of the internal representation of the knowledge base, which enhances processing speed. The result is an inference time of ~50 ms for each profile (*6, 7*).

Evaluation of the expert system

VALAB has been evaluated by analyzing the system's internal validity as well as its expertise.

1. Verification of the internal validity of the knowledge base. This consists of studying the internal logic of a rule group. During the compilation step, the KHEOPS program builds a full representation of the decision tree and explores all linkages between rules to detect and amend any incorrect reasoning. This verification of the internal logic of the program ensures that contradictory situations are not encountered.

2. Evaluation of the system's expertise. This was done by comparing the decision of the expert system with that of another method. The comparison method was defined as the consensus opinion of seven laboratory supervisors (four M.D.s with specialization in chemical pathology, two Ph.D.s in pharmacy with specialization in clinical biochemistry, and one Ph.D. in clinical chemistry). In addition, the decision of each supervisor was compared with the decision of the comparison method. This allowed for an evaluation of VALAB relative to the individual supervisors.

Each supervisor and VALAB separately evaluated 200 final reports chosen at random from abnormal reports already stopped by the mainframe computer. None of the supervisors contributed to the writing of the knowledge base. The supervisors and VALAB were compared with regard to sensitivity, specificity, positive predictive value, and negative predictive value (*8*). A true positive was defined as correctly stopping a wrong report, and a true negative was the validation of a correct report. A false positive was stopping (inappropriately) a correct report, and a false negative was validating (inappropriately) an incorrect report. The expert system is safety-oriented; i.e., sensitivity is highly valued. We prefer rejecting a correct report to accepting an incorrect one.

Results

Use of the preceding result. The delta check was used to obtain decision limits. We used a parameter called the stability coefficient (SC), defined as follows:

$$SC = \text{current value/preceding value}$$

For each analyte, the SC showed a gaussian distribution when results were expressed on a logarithmic scale. Table 1 contains, for each analyte, the 95% confidence interval for the mean SC value. The SC distribution was truncated to yield acceptable decision limits by eliminating the 5% of the values outside the 95% confidence interval and recalculating the 95% confidence interval of the remaining (truncated) population (*9*). This process was reiterated until the 95% confidence interval was stable. The number of truncations performed for each analyte is also shown in Table 1.

To appreciate the performance of the SC, we also compared the SC value, for each test, with other decision parameters currently used in clinical laboratories. These parameters included the theoretical limits for analytical variability given by the French Society for Biology (*10*), day-to-day laboratory precision, Barnett's medically acceptable limits (*11*), or clinically useful limits (*12*). However, these criteria were not incorporated into the expert system. Table 2 shows an example (serum calcium) of these limits as used in the authors' laboratory.

The delta check presented here was unfortunately not performed with regard to the length of time between the present and preceding values. Thus we did not compare our SC value with the rate check proposed by Lacher and Connelly (*13*).

Evaluation procedures. The VALAB decisions were compared with the decisions of each of seven supervisors

Table 1. Acceptability Limits from Stability Coefficients (SC)

	n	95% confidence limits: Before truncation	After 1st truncation	After final truncation	No. of truncations
SC Na	1659	0.936/1.062	0.950/1.046	0.968/1.025	14
SC K	1539	0.739/1.375	0.795/1.273	0.845/1.190	14
SC Cl	1530	0.908/1.091	0.928/1.069	0.957/1.037	11
SC bicarbonate	1658	0.733/1.372	0.791/1.274	0.857/1.176	12
SC protein	1673	0.759/1.295	0.819/1.206	0.882/1.144	14
SC Ca	1048	0.876/1.132	0.902/1.102	0.924/1.075	10
SC urea	881	0.398/2.512	0.532/1.935	0.712/1.374	22
SC creatinine	2118	0.608/1.603	0.721/1.359	0.854/1.178	15
SC glucose	1216	0.421/2.148	0.553/1.675	0.814/1.194	19
SC ionic balance	1526	0.915/1.092	0.930/1.075	0.953/1.054	19

Table 2. Useful Limits for Validation of Serum Calcium Values

Analytical variability (maximum acceptable imprecision):	
Laboratory day-to-day precision	±2.2%
Theoretical (VALTEC) (*9*)	±3.2%
Barnett's medically acceptable limits (*10*)	±4.6%
Clinically useful limits (*11*)	±6%
VALAB coefficient of stability limits	±10%

in terms of sensitivity, specificity, and predictive values. The comparison method was defined as the consensus of the seven supervisors. The results (Table 3) show that VALAB has a sensitivity equal to or greater than that of the human experts. These findings are in agreement with those for other evaluation protocols (*14*).

Operational characteristics. VALAB is located in a microcomputer, connected to the laboratory's mainframe computer. The connection simulates the video display unit used by the supervisors when checking the reports stored in the "to be validated" file on the mainframe. The expert system is fully automated and follows the following steps: (*a*) presentation of the laboratory report on the microcomputer monitor; (*b*) identification of the data in the report and syntax analysis; (*c*) performance of the expert function; (*d*) local display of the decision to accept or reject the reported data (unacceptable data are colored red on the display); (*e*) transmission of the decision to the mainframe computer, which either prints an accepted report or stores a rejected report (along with the reason for rejection); and (*f*) retrieval of the next report. Rejected reports are stored together for convenient evaluation by a laboratory supervisor.

VALAB is used daily for every profile run in the laboratory (e.g., admission profile). It systematically rejects any report containing a test that is not included in the knowledge base. Since November 1988, the VALAB system has been in routine operation, handling 22 biochemical tests (sodium, chloride, potassium, CO_2, total protein, calcium, urea, creatinine, glucose, ionic balance, phosphate, iron, uric acid, cholesterol, triglycerides, total bilirubin, alkaline phosphatase, γ-glutamyltransferase, alanine aminotransferase, aspartate aminotransferase, lactate dehydrogenase, and creatine kinase). The system has automatically validated >50 000 reports. At this time, the VALAB system can process about 70% of the total work load of the laboratory; it validates >50% of the reports.

VALAB provides an expertise for the routine work that runs silently in the background of the laboratory organization. The system requires only ~50 ms to validate a report and eliminates most of the human–computer interface that is normally required for routine validation of data.

Discussion

To date, there have been few evaluations of operational expert systems in the clinical laboratory. In the

Table 3. Evaluation of the Expert System VALAB

	T−	T+	F−	F+	Accepted	Rejected	Sensitivity	Specificity	+PV	−PV
VALAB	104	27	9	60	113	87	0.750	0.634	0.310	0.920
Expert consensus	164	36	0	0	164	36	1	1	1	1
Expert 1	142	29	7	22	149	51	0.806	0.866	0.569	0.953
Expert 2	162	8	28	2	190	10	0.222	0.988	0.800	0.853
Expert 3	157	12	24	7	181	19	0.333	0.957	0.632	0.867
Expert 4	160	23	13	4	173	27	0.639	0.976	0.852	0.925
Expert 5	148	20	16	16	164	36	0.556	0.902	0.556	0.902
Expert 6	154	25	11	10	165	35	0.694	0.939	0.714	0.933
Expert 7	143	18	18	21	161	39	0.500	0.872	0.462	0.888

T−, true negative; T+, true positive; F−, false negative; F+, false positive; −PV, negative predictive value; +PV, positive predictive value.

past, these systems have been hindered by scientific and ethical problems as well as difficulties in integrating the systems into the laboratory environment.

Scientific Considerations

The inference engine we used (KHEOPS) meets most of the expected requirements of such a process as proposed by Winkel (*1*). The system is linked to the local laboratory information system as an intelligent peripheral device, carries its own graphics and statistical programs, and has a large computational capacity.

The structure of the knowledge base must meet certain criteria (see *Methods*) related to the maintenance, modularity, and optimization of the expert system.

Our application appears to satisfy user-interface requirements. The participation of computer scientists was necessary to help produce the compilation of rule bases (which helped ensure consistency within the knowledge base). The result of this collaborative effort was a new "concentrated" program written in the C programming language. In the run-time version of this program, the rules are inaccessible.

The production rules were written by a medical doctor (E.R.) skilled in information science, who used the knowledge and the experience of our medical staff as well as the literature (*15*). The internal logic of the system cannot be modified by laboratory personnel. However, most of the limiting ranges used in this study can be easily changed to permit implementation in different laboratories. In addition, the individual clinical chemist can personalize the system to include his or her choice of titles, units, acceptability ranges, and the presence or absence of delta-check procedures.

A few drawbacks of the system should be pointed out. The most important is that clinical information is not incorporated within the rules of the system. Unfortunately, it is not yet possible to automatically access this clinical information because the hospital network in the institution where our study was performed is still under development. Eventually, this network will allow clinical practitioners to input the clinical diagnosis when they complete the laboratory requisition. When this feature becomes available, we expect to significantly enhance the VALAB system.

Another drawback of the system is its inability, during the delta check, to consider the time elapsed since the preceding result; thus we were unable to adapt to our system the rate check proposed by Lacher and Connelly (*13*). Unfortunately, we do not yet know how to manage this important item.

A final problem is that not all the activities of the laboratory are covered by VALAB, e.g., stat analyses and hematologic and immunologic analyses. We are currently working on incorporating these other aspects of laboratory analysis into the system.

Ethical Considerations

The ethical problem created by using automated processing of biochemical data remains unclear. This system is less dramatic than are the expert systems used in clinical diagnosis. Nevertheless, we believe that the system should not be used as the sole decision maker but rather as an adjunct to the supervisor. Additional opinions regarding this topic will be obtained from other hospital laboratories participating in a multicenter evaluation of the expert system.

Integration of VALAB into the Laboratory Environment

The integration of VALAB within the laboratory has been relatively easy. Most of the requirements were known at the beginning of the project. Use of a prototype permitted us to develop a flexible tool incorporating complicated rule bases and a sophisticated software package.

The VALAB expert system has had a major impact on the laboratory. It has led to improved test turnaround times (*16*) and has provided relief to the evaluators in charge of daily validation of data. This tedious, iterative, and important work was reduced by about 75%, with an increase in the confidence of laboratory results. In addition, VALAB has helped focus the attention of laboratory supervisors on the reports that the expert system rejects. These are the reports most likely to be erroneous.

We have designed a prototype expert system that helps hasten the delivery of biochemical laboratory data and relieves the team in charge of validating laboratory reports. This system, VALAB, is now evolving in many directions. For example, we are developing knowledge bases for handling other disciplines (e.g., blood cell counting and leukocyte differential, blood proteins). In addition, we are improving quality control of the system by providing printouts of reports that can be checked manually against the automated decision made by the expert system. Finally, the program is being distributed commercially in the following three formats: (*a*) as a dedicated work station or stand-alone expert system connected to a mainframe computer in the laboratory; (*b*) as a system that is fully integrated within the software of the laboratory information system (creating a system for validation that is invisible to the user); and (*c*) as a decentralized application in which VALAB is used at the bench. In this last case, the expert system can be used to monitor large equipment and to process all quality-control data.

This work was supported by grants from the Ministère de la Recherche (Paris) and from the Région Midi-Pyrénées (Toulouse). The valuable assistance of Dr. M. Ghallab (Toulouse) for scientific advice and of Drs. Peter Wilding and Michael Astion (Philadelphia) for criticism and English revision is gratefully acknowledged.

References

1. Winkel P. The application of expert systems in the clinical laboratory. Clin Chem 1989;35:1595–600.
2. Winston PH. Intelligence artificielle. Paris: Interédition, 1988.
3. Rogari E, Philippe H, De Graeve J, Valdiguié PM. Le système expert "VALAB" au laboratoire de biochimie: validation assistée par ordinateur (VAO). Innov Technol Biol Med 1990;11:75–88.
4. Fieschi M. Intelligence artificielle en médecine des systèmes experts. Paris: Masson, 1984.

5. Nosanchuk JS, Gottman AW. Cusums and delta checks. Am J Clin Pathol 1974;62:707–15.
6. Ghallab M. Compilation de bases de connaissances. In: Pastre D, ed. Intelligence artificielle. Actes des journées nationales de Toulouse, 14–15 March 1988. Paris: Teknea, 1988:231–53.
7. Ayel M, Chein M, Pipard E, Rousset MC. De la cohérence dans les bases de connaissances. *Ibid.* 357–88.
8. Rumeau-Rouquette C, Bréart G, Padieu R. Méthodes en épidémiologie. Paris: Flammarion Médecine Sciences, 1986.
9. Whitehead TP, Browning DM. The U.K. national quality control scheme. In: Rappoport AE, ed. Quality control in clinical chemistry. Bern: Hans Huber Publishers, 1972:30–47.
10. Vassault A, Grafmeyer D, Naudin C, et al. Protocol for the validation of methods. Ann Biol Clin (Paris) 1986;44:686–745.
11. Barnett RN. Clinical laboratory statistics. Boston: Little, Brown and Co., 1971.
12. Schoen I, Graham GA, Cuter E, Bandi Z, Georgeswill D. Clinically useful limits (CUL) criteria as a basis for quality control including minimal and optimal goals for quality control. Am J Clin Pathol 1989;91:171–81.
13. Lacher DA, Connelly DP. Rate and delta checks compared for selected chemistry tests. Clin Chem 1988;34:1966–70.
14. Quaglini S, Stefanelli M. A performance evaluation of the expert system ANEMIA. Comp Biomed Res 1988;21:307–23.
15. Siest G, Henny J, Schiele F. Interprétation des examens de laboratoire. Valeurs de référence et variations biologiques. Paris: Copedith (and Basel: Karger), 1981.
16. Hilborne LH, Oye RK, McArdle JE, Repinski JA, Rodgerson DO. Use of specimen turnaround time as a component of laboratory quality. Am J Clin Pathol 1989;92:613–8.