Eur J Clin Chem Clin Biochem 1996; 34:371–376 © 1996 by Walter de Gruyter · Berlin · New York

The Performance of the Knowledge-Based System VALAB Revisited: An Evaluation after Five Years[1])

Pierre M. Valdiguié[1], *Eric Rogari*[2], *Joël X. Corberand*[3] and *Bernard Boneu*[3]

[1] Laboratory of Chemical Pathology, Rangueil University Hospital Toulouse, France
[2] Erems Company, Flourens, France
[3] Laboratory of Haematology, Rangueil University Hospital, Toulouse, France

Summary: In 1988, inundated by the tedious work of validation of laboratory reports in a large hospital biochemistry laboratory, we designed VALAB, a knowledge-based system specially dedicated to this iterative function.

Coping at first with a few biochemical tests, the program has been progressively expanded to forty-five common chemical tests. Simultaneously some new rules have been introduced to "weight" the conclusion in different circumstances and rules taking into consideration some clinical data have also been written.

Moreover the program moved to other disciplines, pH and blood gases, haematology and coagulation. Accordingly the evaluation protocol has been modified, incorporating a new step, the consensus decision of the pathologists, operating within the initial protocol and based upon the various criteria of epidemiology.

These major changes and improvements have led us to check and describe again the performance of this updated VALAB knowledge-based system.

Introduction

In large hospital laboratories that use high throughput equipment, the task for human validation of final reports is very important, in spite of the help provided by efficient laboratory information systems. It is time consuming and highly dependent on the skill and experience of the supervisors. Therefore we decided in 1988 to use "artificial intelligence" and to carry out a knowledge-based system project to aid decision making and to perform an automated validation of data. The program was first designed for an electrolyte profile (1) but it has been rapidly expanded to handle 22 tests commonly run in the clinical chemistry laboratory (2). Right now the system is able to deal with 45 commonly used tests. Simultaneously, new rules have been added to cope with clinical data, the final decision is improved by "weighing" rules that are used in different clinical circumstances. Moreover, in addition to its use in the Chemical Pathology laboratory, the system has also been allocated to other disciplines of laboratory medicine, Haematology (3, 4) and Haemostaseology, where automated equipment is also operated. When this occurred, the first evaluation protocol (2) was modified and accordingly also changed in Clinical Chemistry.

Since many amendments and improvements have been introduced in the program, we have thought it would be interesting to check again and report the performance in the three disciplines of this updated version of our knowledge-based system VALAB.

Material and Methods

Material

The knowledge-based system operates on a microcomputer IBM-compatible PC (Compaq, Microdis, 31700 Blagnac, France) containing an Intel 80386 or 80486 processor, 4 megabytes of RAM, a 80-megabyte hard disk and Hercules or VGA graphics.

The software runs under MS-DOS and uses the generator (inference engine) KHEOPS (5) from the Laboratoire d'Automatique et d'Analyse des Systèmes, an institute of the Centre National de la Recherche Scientifique in France. KHEOPS uses forward chaining as the reasoning process that is applied to the knowledge base represented in the form of production rules. It is moreover able to compile the rule base.

Methods

1. The various tests included in the knowledge base are listed in tables 1–2, covering Biochemistry, Haematology and Coagulation.

2. The production rules (more than 20 000) represent the knowledge and are expressed in conditional (if-then) form. There are four sets or rules:

(a) The ones representing the core of the system are devoted to the various criteria selected to help decide whether to validate laboratory data. VALAB actually uses the following information for every patient data: acceptable limits, internal coherence between analyte results which are physiologically related, delta check, origin of the sample, i.e. identification of the ward and the medical

[1]) This work was supported by grants from the Conseil Régional de Midi Pyrénées and an award from the Société Française de Biologie Clinique.

speciality, "Stat" analysis or not, out or in-patient, age, sex, comments on the sample quality.

(b) Some rules (weighing rules) define dynamically acceptability thresholds for each patient, as various trends for that patient are noticed. This "qualitative reasoning" approach (5, 6) is a characteristics of the second generation of knowledge-based systems. There are weighing rules for instance to modify the acceptable ranges in various analyte values or in delta check acceptance.

(c) For each test some "negative" rules have been written in order to restrict the validation of a normal value that would not be in accordance with other data.

(d) particular rules cope with clinical or therapeutical data. They are of major interest and must be developed in the future. An example of these different rules is given in the Appendix.

All these rules are divided into 100 rule groups, each rule group containing between 100 and 300 elementary rules which are related

Tab. 1 List of the chemical tests expertised by VALAB.

General analyses	Specialised analyses	pH and blood gases
Sodium	Glycated haemoglobin	pH
Potassium	Fructosamine	pO_2
Chloride	Iron	pCO_2
Bicarbonate	Ferritin	Bicarbonate
Total protein	Transferrin	Standard bicarbonate
Anion gap	Coefficient of iron saturation	Total CO_2
Delta Na-Cl	Alkaline phosphatase	Base excess
Urea	γ-Glutamyl transferase	O_2 Saturation
Creatinine	Amylase	Haemoglobin
Glucose	Lipase	Total O_2
Uric acid	Total bilirubin	
Cholesterol	Conjugated bilirubin	
HDL cholesterol	Unconjugated bilirubin	
LDL cholesterol	Alanine aminotransferase	
Apolipoprotein A1	Aspartate aminotransferase	
Apolipoprotein B	Lactate dehydrogenase	
Triacylglycerols	Creatine kinase	
Calcium	Creatine kinase MB	
Phosphate	C-Reactive protein	
Plasma Mg	α_1-Acid glycoprotein	
Erythrocyte Mg	Haptoglobin	
	Free T_3	
	Free T_4	
	Thyrotropin	

Tab. 2 List of the tests known by the expert system in haematology and haemostaseology. Numerous other data, clinical, therapeutic, morphological (erythrocytes) or plasma aspect (turbidity, haemolysis) are also taken into consideration.

Cellular haematology	Haemostaseology
Haemoglobin	*Quick* time (QT) or prothrombin time (PT)
Mean Corpuscular Volume	International Normalized Ratio
Mean Corpuscular Haemoglobin	Factor V
Mean Corpuscular Haemoglobin Concentration	Factor VII + X
Erythrocytes	Factor II
Platelet Cell Volume	Fibrinogen
Reticulocytes	Heparin level (unfractionated heparin)
Erythrocytes morphology	Heparin level (low molecular mass heparin)
Platelets	Activated partial thromboplastin time (APTT)
Leukocytes	Thrombin clotting time (TCT)
Neutrophils	Thrombin clotting time corrected by protamin sulphate (TCTPS)
Eosinophils	Reptilase time (RT)
Basophils	Factor VIII
Lymphocytes	Factor IX
Monocytes	Factor XI
Immature granulocytes	Factor XII
Promyelocytes	Bleeding time (BT)
Myelocytes	Ethanol test
Metamyelocytes	Circulating anti coagulant CAC
Plasma cells	
Atypical lymphocytes	
Erythroblasts	
Leukoblasts	
Lymphome cells	
Erythrocyte Sedimentation Rate (1 h)	

to a similar topic. An example of the strategic path used, e. g. for the validation of a high aspartate aminotransferase value, is shown in the Appendix.

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. This "pretreatment" of the internal representation of the knowledge base results in a total inference time, which varies for each report according to the number of data, but does not exceed approximately 500 ms.

3. The evaluation protocol has been modified in order to introduce a new step: the clinical chemists' and pathologists' consensus that is the reference decision, with which either VALAB or every supervisor decision will be compared. Four MDs with specialisation in Chemical Pathology, two PhDs in pharmacy with specialisation in Clinical Biochemistry and one PhD in Clinical Chemistry for the laboratory of Clinical Chemistry and three MDs specialised in Haematology for the Laboratory of Cellular Haematology and four Clinical Pathologists in Haemostaseology were the human referees.

They had to check separately 338 patient reports in chemistry, 384 for pH and blood gases, 357 in haematology and 550 in haemostaseology. The control process was conducted along the epidemiological method used to assess the sensitivity, specificity and predictive values of a clinical symptom or a laboratory test. These values can be calculated and compared between the knowledge-based system and the human observers.

T(+) or true positive is defined as correctly stopping a wrong report, T(−) or true negative is the validation of a correct report. F(+) or false positive is the inappropriate rejection of a good report and F(−) is badly accepting an incorrect report.

Review of the formulae shows that the emphasis must be largely given to sensitivity and negative predictive value because they both contain the unacceptable F(−).

Sensitivity
(proportion of rejected incoherent reports) $\frac{T(-)}{T(+) + F(-)}$

Specificity
(proportion of accepted coherent reports) $\frac{T(-)}{T(-) + F(-)}$

Positive predictive value
(proportion of incoherent reports within the rejected ones) $\frac{T(+)}{T(+) + F(+)}$

Negative predictive value
(proportion of coherent reports within the accepted ones) $\frac{T(-)}{T(-) + F(-)}$

Furthermore, the system has also been submitted to a national multicentric evaluation in five different laboratories, with 4 large hospital laboratories of clinical chemistry and 1 big private laboratory of clinical pathology, representing a total of 19 referees.

4. Statistical data are available concerning the activity and the performance of VALAB within the various laboratories of our hospital, with emphasis on results that are considered invalid, and which must be viewed by the user along with the reasons given by the system for the rejection.

Results

Data from the evaluation protocols and from the statistical activities are presented here.

1. Evaluation results

1.1 In Clinical Chemistry

In this study 338 reports were included. The VALAB decisions on the one hand and the human ones on the other were both compared to the collegial decision, defined as the consensus of the various supervisors. Data are presented in table 3.

The various steps were

(a) to check first the 338 reports within a single period of time for each of the seven supervisors in order to consider the tiring effect of such a batch of results to be validated. Fifty seven reports showed discrepancies between the various supervisors and needed a search for consensus which was easily met.

(b) Taking into account this consensus decision, two reports accepted by VALAB but previously blocked by the medical staff were thus accepted and therefore 2 F(−) moved to 2 T(+); and twenty nine reports validated by the staff but firstly rejected by VALAB were accepted by the system after some modifications in the "weighing" rules, resulting in 29 F(+) becoming 29 T(+).

(c) To improve the system performance again, we adjusted some upper limits and accordingly the VALAB final decision was to reject four reports that were previously accepted, 4 F(−) becoming 4 T(+), and to accept nine cases rejected before the correction, 9 F(+) moving to 9 T(−).

With these last figures, sensitivity, specificity and predictive values were calculated again, showing a sensitivity of 100% and a negative predictive value of 100%; these are the main values to consider because there is F(−) in their definition and we cannot accept a system that inappropriately validates a wrong report.

1.2 Multicentric evaluation in Clinical Biochemistry

Data were collected under the same conditions within the various selected laboratories at the national level. VALAB was connected to different Laboratory Information Systems and 1675 reports were examined.

The general conclusions are presented in table 4. In four laboratories 38.5% of the reports accepted by the medical staff were also validated by VALAB, except in one hospital laboratory dealing only with emergency testing for very severe diseases and without previous results, where the knowledge-based system accepted only 5% of the 65% validated by the staff.

1.3 In haematology

The evaluation was performed by three clinical pathologists on 357 reports randomly selected from the file of reports needing a medical validation.

As in the clinical chemistry protocol, we performed the first individual validation with the fatigue effect for human observation.

Some reports were then modified after consensus, finally producing VALAB's validation after amendment

of some parameters for best fit with the pathologists' consensus, which is considered as the ideal decision.

Between the first two steps, 89 reports produced variable decisions amongst the three pathologists, necessitating a consensus that was met easily, except for 5 reports which were therefore withdrawn.

All these data are presented in table 5.

1.4 In haemostaseology

The evaluation was performed by four MDs who compared their decision for 550 reports with that of the VALAB. The same protocol was again used and gave the following results:

111 reports needed a consensus, 94 lacking agreement from the four pathologists, 17 being blocked by VALAB and accepted by the medical staff (F(+)). There was no F(−) in the expert system analysis.

After consensus decision and modification of some parameters and some weighing data in VALAB's program, the final calculation was excellent and gave 1.00 for sensitivity and negative predictive value, with no residual F(−).

1.5 pH and blood gases

The protocol covered 384 reports. During the first VALAB's run we noted 7 F(−) and 71 F(+) most of them, 51, due to a very high pO_2 caused by oxygenotherapy. After the consensus meeting, the acceptable limits for pO_2 were modified and the 7 F(−) became 7 T(−). We decided also to ask the intensive care units to mention the oxygen therapy on the request forms, this

Tab. 3 Epidemiological data for the three step evaluation of the expert system VALAB in clinical biochemistry.
(a) is the primary comparison between VALAB and seven supervisors.
(b) is the result obtained after consensus.
(c) is the final decision of VALAB after modification of some parameters taking into consideration the consensus decision.

	T(+)	T(−)	F(+)	F(−)	Accepted	Rejected	Sensitivity	Specificity	(+)PV	(−)PV
VALAB (a)	127	165	39	7	172	166	0.947	0.808	0.765	0.959
Staff (a)	57 reports among 338 needed a consensus between the seven supervisors									
VALAB (b)	157	167	10	4	171	167	0.975	0.944	0.940	0.977
Staff (b)	132	175	29	2	204	134	0.820	0.989	0.985	0.858
VALAB (c)	161	176	1	0	176	162	1.000	0.994	0.994	1.000

Tab. 4 Average of the data from 19 observers and from VALAB in a multicentric national evaluation for the clinical chemistry program.

	Sensitivity	Specificity	Positive PV	Negative PV
Human data				
Mean of the 19 human observers	82.8	92.8	75.3	94.7
Range	62.2− 93.2	71.9−98.5	38.9−97.5	87.0− 98.9
Expert system data				
Mean of VALAB's data within the 5 locations	98.1	31	27.2	97.3
Range	95.4−100	51.0−44.2	6.1−47.3	92.0−100

Tab. 5 Evaluation protocol in haematologic cytology.
(a) is the primary comparison between VALAB and three supervisors.
(b) is the result obtained after consensus.
(c) is the final decision of VALAB after modification of some parameters taking into consideration the consensus decision.

	N	T(+)	T(−)	F(+)	F(−)	Accept	Reject	Sensitivity	Specificity	(+)PV	(−)PV
VALAB (a)	357	22	267	49	19	286	71	0.537	0.845	0.310	0.934
Staff No. 1 (a)	357	29	293	23	12	305	52	0.707	0.927	0.558	0.961
Staff No. 2 (a)	357	24	299	17	17	316	31	0.586	0.946	0.586	0.946
Staff No. 3 (a)	357	40	297	19	1	298	59	0.976	0.940	0.678	0.997
VALAB (b)	352	26	261	42	23	284	68	0.531	0.861	0.382	0.919
Staff No. 1 (b)	352	19	272	31	30	302	50	0.388	0.898	0.380	0.901
Staff No. 2 (b)	352	26	289	14	23	312	40	0.531	0.954	0.650	0.926
Staff No. 3 (b)	352	34	282	21	15	297	55	0.694	0.931	0.618	0.949
VALAB (c)	352	41	254	49	8	262	90	0.837	0.838	0.456	0.969

information being therefore taken into consideration by the knowledge-based system.

2. Statistical data concerning routine operation

The three laboratories are using VALAB for a round the clock service.

The reports submitted to the expert system are not identical, and they vary according to the discipline.

In Chemical Pathology the system examines only the reports already blocked for any abnormality by the laboratory information system and then stored in a special file of reports to be validated. VALAB regularly explores this file and, according to its knowledge, either rejects or validates the reports, which are, in this case, immediately sent through the hospital network and printed out. The remaining reports, with indication of the reason for VALAB's rejection, are reviewed on the screen of the laboratory information system by the medical staff.

In Haematology and Haemostaseology, entire reports may be considered as abnormal by the laboratory information system and thus VALAB has to expertise all the data stored in the file.

An example of the activity of the knowledge-based system during a relative quiet fortnight of July 1995 is given in table 6.

Discussion

VALAB can be considered as a screening program dedicated to the automated selection of reports needing a human view, in order to either accept them as valid or have them rerun or, mainly in Haematology or Haemostaseology, have them checked comprehensively with dialogue with the physician.

It was most important, of course, to perform a very strict evaluation to check the adequateness of VALAB expertise before the routine implementation of such an automated process. The method used for the evaluation is derived from the epidemiological protocols. It gave satisfactory results after addition of the consensus step, which represents ideal decision from the medical point of view.

We did not strictly follow *Miller*'s proposal (8) who distinguished three levels of evaluation: evaluation of research contribution, validation of knowledge and performance, evaluation of the clinical efficacy of the operational system, because we limited our protocol to steps 2 and 3.

Actually, VALAB is not a clinical system to be used by physicians for interpretation of laboratory data or support for diagnosis. It is rather a tool for senior clinical chemists or pathologists remaining within the laboratory.

The only data available for evaluation of knowledge-based systems are clinical data for the performance of knowledge-based systems in their support of the interpretation of laboratory findings (9). The strategy used by *Wyatt* (10) is to answer the following questions:

i) is the system wanted and of good quality? (structure),

ii) is the system pleasant to use and does it reason appropriately? (reasoning process),

iii) does it say sensible things and draw valuable conclusions? (outcome); and the means of attaining this goal are peer review and field trials.

We may consider that we have attained these objectives, because VALAB is now spread over 35 European laboratories, and because in our hospital, since 1988, we have never had any question or argument from the clinicians related to the patient reports validated by the knowledge-based systems.

VALAB has now incorporated second generation concepts (6, 7) and is able to weigh its decision according to various predefined items.

The conditions of operation can be selected within the main frame computer (laboratory information system) to which the knowledge-based system is connected as an analytical instrument; it can be as to examine either only pathological reports or any report if the limits of normality are strictly narrowed.

VALAB has been designed as a tool for helping in the tedious and iterative process of final medical validation, and all the laboratories in Europe equipped with this decision support program are using it for this task in the clinical chemists' or pathologists' office. However, it is obvious that many laboratories are limiting their validation at the bench, where they perform sophisticated pro-

Tab. 6 Total number of reports submitted to VALAB for 2 weeks in July 1995.

	Reports seen by the expert system	Reports effectively expertised	Reports validated	Fraction of validation (%)
Chemical Pathology	3378	3198	1625	51.3
Haematology	4063	3664	2788	76.1
Haemostaseology	2490	2415	2107	87.2

cess, using quality control, delta-check, mean of normals appreciation as part of the technical validation. It is therefore interesting to consider whether VALAB cannot move to the bench, become embedded in the advanced instrument workstation, and interface between high throughput equipment and laboratory computer. Such a development would seem imminent, particularly within the "Openlabs" project of the European Community (11, 12).

Whatever the location of VALAB within the laboratory, one advantage must be emphasised, i.e. the improvement of turn around time due to a rapid check and often validation of abnormal reports without waiting for a human decision. The application to various disciplines where automated equipment provides a high volume of data should also be mentioned, the program for immunoanalysis being presently under development.

Concerning the ethical problem, we have to remember that VALAB is an aid to the decision maker, and is not intended to supplant him (her); actually it represents a cooperative effort of man and machine (13).

Acknowledgements

The valuable help and cooperation of the members of the medical staff in the three laboratories is gratefully acknowledged.

Appendix

1. Examples of the different rules

a) *Basic production rule* (haemoglobin)
If there is a low value for haemoglobin.
If the patient is located within a surgical intensive care unit,
Then decrease the acceptability of this low haemoglobin by 30 g/l.

b) *Weighing correlation rule* (serum calcium)
If there is a low value for calcaemia,
If there is a result for serum creatinine,
If the creatininaemia is higher than 150/300/500 µmol/l,
Then increase the acceptability of this low calcaemia by −0.1/−0.2/−0.3 mmol/l.

c) *Negative rule* (*Quick* time)
If there is an increase of *Quick* time higher than 8 seconds,
If there is a result for activated partial thromboplastin time,
If the increase of activated partial thromboplastin time is lower than 3 seconds,
Then it is not possible to validate such a value for *Quick* time.

d) *Clinical rule* (pO_2 in blood gases)
If there is any oxygen therapy,
If there is a result of pO_2 higher than 100 mm Hg
Then it is possible to validate such an abnormal value of pO_2.

2. Example of strategic reasoning pathway

If there is a very high value for serum aspartate aminotransferase (e.g.; > 300 IU/l, 37 °C):
- Look for other data able to justify this value:
  - Myocardial infarction context:
    High or very high creatine kinase-MB, creatine kinase, myoglobin, cardiologic intensive care unit location, clinical information on myocardial infarction.
  - Or hepatitis context:
    Very high serum alanine aminotransferase, high or very high conjugated bilirubin, infectious disease, high C-reactive protein, digestive diseases ward location, clinical information on acute hepatitis.
  - Or other context concerning liver or pancreatic disease.
  - Or chemotherapy context.
- Control that there is no negative rule triggered to forbid the acceptability of such a value of serum aspartate aminotransferase:
  - e.g. very low result for serum alanine aminotransferase.

References

1. Rogari E, Philippe H, De Graeve JS, Valdiguié PM. Le système expert « VALAB » au laboratoire de biochimie: validation assistée par ordinateur (VAO). Innov Technol Biol Med 1990; 11:75–88.
2. Valdiguié PM, Rogari E, Philippe H. VALAB: expert system for validation of biochemical data. Clin Chem 1992; 38:83–7.
3. Corberand JX, Rogari E, Valdiguié PM. A computer-assisted validation system for laboratory data in haematology: "Valab/Haemato". Ann Biol Clin 1993; 51:228–30.
4. Corberand JX. Computer-assisted validation of hematologic reports. Lab Medica International 1994; July–August:16–20.
5. Ghallab M, Philippe H. A compiler for real-time knowledge base systems. In: Proceedings of the International Workshop on Artificial Intelligence for Industrial Applications. Hitachi, Japan, 1988, 387–93.
6. Winkel P. The application of expert systems in the clinical laboratory. Clin Chem 1989; 35:1595–1600.
7. Pohl B, Beringer C, Walther S, Melzer J, Burow F, Schmidt-Schauss M, et al. Neue Verfahren zur Erstellung wissenbasierter Befundungssysteme mit der Expertensystemschale Pro. M.D. Lab Med 1994; 18:577–81.
8. Miller PL. The evaluation of artificial intelligence systems in medicine. Comput Methods Programs Biomed 1986; 22:5–11.
9. Quaglini S, Stefanelli M. A performance evaluation of the expert system ANEMIA. Comput Biomed Research 1988; 21:307–23.
10. Wyatt J, Spiegelhalter D. Evaluating medical expert systems: what to test and how? Med Inform 1990; 15:205–17.
11. De Graeve JS, Cambus JP, Gruson A, Valdiguié PM. Automated technical validation. A real time algorithm for decision support. Proceedings of the Xth International Conference on Computing in Clinical Laboratories, Jerusalem, May 29–Jun 2, 1994. Clin Chim Acta. In press.
12. Groth T. Openlabs advanced instrument workstation services. Proceedings of the Xth International Conference on Computing in Clinical Laboratories: 1994 May 29–Jun 2; Jerusalem. Clin Chim Acta. In press.
13. Catrou PG. Clinical laboratory informations, the promised land. Am J Clin Pathol 1995; 103:677–78.

Received September 8/December 6, 1995

Corresponding author: Prof. Dr. Méd. Pierre M. Valdiguié,
Laboratoire de Biochimie, Hôpital universitaire de Rangueil,
1 Avenue J. Poulhès, F-31054 Toulouse cédex, France