Eur J Clin Chem Clin Biochem 1996; 34:371–376

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

Eur J Clin Chem Clin Biochem 1996; 34:377–384 © 1996 by Walter de Gruyter · Berlin · New York

TECHNICAL REPORT

A Multicentre Evaluation of Tumour Marker Determinations Using the Automatic Enzymun-Test® Systems ES 300 and ES 600/700

Arie van Dalen[1] and *Anne-Christine Kessler*[2]

[1] Department of Nuclear Medicine, Groene Hart Ziekenhuis, Bleuland location, Gouda, The Netherlands
[2] Research Centre Tutzing, Boehringer Mannheim, Tutzing, Germany

Summary: A multicentre evaluation of the determination of carcinoembryonic antigen (CEA), the cancer antigens CA 15-3, CA 19-9, CA 72-4 and CA 125 (II generation), the cytokeratin 19 marker Cyfra 21-1 and α-foetoprotein (AFP) using the Enzymun-Test® System (ES 300 and ES 600/700) was performed in 23 laboratories. The tumour markers were measured in a total of 4266 human serum samples.

The intra-assay precision was less than 5% in 80% of all serum samples investigated and in 95% of the serum samples at or above the cut-off level of the tumour markers. Inter-assay precision was less than 10% in 86% of the marker determinations. The interlaboratory survey also showed high reproducibility for the determination of all the tumour markers. In 3 laboratories the results of CA 15-3 in 283 serum samples were compared with the IRMA method of CIS bio international. The regression coefficient, r, was 0.967. In 4 laboratories the results of CEA in 312 samples were compared with the results obtained on the IMx analyser. The regression coefficient, r, was 0.967. In benign gynaecological diseases, CA 125 (II) was most frequently elevated in endometriosis. In gastrointestinal diseases it was proven that CEA is still the marker with the highest sensitivity as compared with CA 19-9 and CA 72-2 (59% with healthy controls as the reference group and 44% with patients having benign gastrointestinal disease as the control group). In pancreatic cancer CA 19-9 showed the highest sensitivity (78% and 62% respectively). In gastric cancer the three markers did not show statistically different results. When the gastric cancer patients were divided according to stage, CA 72-4 appeared to be more sensitive than CA 19-9 only in stage IV.

Introduction

The development of immunochemistry analysers made possible the automated determination of tumour markers. Recently, developments in immunoassay methodology were reviewed (1). However, at present, the results of assays from some manufacturers for an individual tumour marker vary considerably, as shown by external quality assurance schemes (EQAS) in different countries in Europe (2–4). The analytical performance of ES 300 and ES 600 systems in smaller pilot studies has been reported (5, 6).

In this study we compare results for the determination of carcinoembryonic antigen (CEA), the cancer antigens CA 15-3, CA 19-9, CA 724 and CA 125 (II), Cyfra 21-1 (a marker of cytokeratin 19) and α-foetoprotein (AFP) on the ES 300 and ES 600/700 analysers, in a study involving 23 laboratories. In addition to assessing the technical performance of the analysers, comparative studies were carried out with routinely used assays, which were also performed manually or on other types of automated apparatus. Special attention was given to the second generation of the CA 125 determination and to the application of the new marker CA 72-4 in gastrointestinal cancer. We also investigated differences in cut-off levels based on normals and on patients with relevant benign diseases.

Material and Methods

Samples

The multicentre evaluation was performed in 23 laboratories. The tumour markers were measured in 4266 human serum samples. These samples were obtained from local patients (N = 2170) during routine investigations in the institutes;

healthy subjects	N = 1098;
benign pulmonary diseases	N = 135;
benign gastrointestinal diseases	N = 103;
benign liver disease	N = 85;
colon carcinoma at diagnosis	N = 226;
gastric carcinoma at diagnosis	N = 110;
pancreatic carcinoma at diagnosis	N = 63;
ovarian carcinoma patients	N = 158 and
benign gynaecological diseases	N = 118.

Assays

Enzymun-Test® CEA, CA 15-3, CA 19-9, CA 72-4, CA 125 II, Cyfra 21-1 and AFP (Boehringer Mannheim Diagnostics, Germany).

IMx system CEA, CA 19-9, CA 125 and AFP (Abbott Laboratories, USA).

Cobas Core EIA CA 19-9, CA 125 II and CEA (Hoffmann La Roche, Switzerland).

AIA 1200 CEA and AFP (Tosoh Corporation, USA).

ENZELSA CA 15-3, CA 19-9, CA 125 II and Cyfra 21-1 (CIS bio international, France).

Stratus CEA (Baxter Diagnostics Inc., USA).

IRMA CA 19-9 and CA 72-4 (Centocor Diagnostics, USA).

All kits were taken from current production batches. The assays were performed according to the manufacturer's instruction. All laboratories were equipped with the fully automated Enzym-Test® System ES 300 or ES 600/700.

Evaluation participants

France:

L. Deneux, Institut Curie, Paris

Italy:

G. Aimo, Ospedale Molinette, Torino
F. Aregnati, Ospedale Riuniti, Jesi
S. Belloli, Universita degli Studi di Brescia, Brescia
A. Lofaro, Ospedale Civile "Lotti", Pisa
M. Onetto, Lab. di Patologia Clinica, Genova
P. Piantino, Ospedale Molinette, Torino
G. Ruggeri, Universita degli Studi di Brescia, Brescia
W. Taccone, Lab. Analisi Centrale, Avellino
G. Vignati, Ospedale Civile, Magenta

The Netherlands:

H. W. A. de Bruijn, Academisch Ziekenhuis, Groningen
A. van Dalen, Groene Hart Ziekenhuis, Gouda
J. Marrink, Academisch Ziekenhuis, Groningen

Portugal:

A. Carvalho dos Santos, Hospital Santo Antonio, Porto
M. Pereira de Melo, Hospital Santa Maria, Lisbon

Spain:

J. Alarco, Hospital Virgen del Pino, Las Palmas
C. Blanco Barros, Hospital de Getafe, Madrid
C. Cacero, Hospital Materno Infantil, Malaga
C. Conzales Valverde, Hospital Xeral de Vigo, Vigo
M. C. Pastor, Hospital HNOS, Trias i Pujol, Badalona

Switzerland:

P. Borer and *R. Keller*, Klinik Barmelweid, Barmelweid
M. Caiezel, Kantonspital, Zug
H. Zihlmann, Kantonales Spital, Sursee.

Control material

Tumour marker controls level 1 and 2 produced by Boehringer Mannheim Diagnostics. Interlaboratory study samples TM 2/92 (A and B) of the German Society for Clinical Chemistry (Bonn, Germany).

Evaluation protocol and statistical methods

All participants followed the protocol provided by Boehringer Mannheim Research Centre.

Intra-assay precision

Native patient samples were analysed 10 times in one series. The mean ($\bar{x}$), the standard deviation (S. D.) and the coefficient of variation (CV%) were calculated.

Inter-assay precision

The tumour marker controls were analysed as single determinations in 3–5 different series. The mean, S. D. and CV% were calculated.

Interlaboratory survey

The samples TM 2/92 A and B were determined as one single determination. The medium and scatter (given as the 16th and 84th percentiles) of both samples A and B of all results were calculated for each analyte.

Comparison studies

Patient samples were measured using the Boehringer Mannheim enzyme immunoassay in parallel with the routine method of each laboratory. Correlation was calculated using the method of *Passing & Bablok* (7).

Reference values

Sera from healthy subjects (blood donors, hospital staff) were assayed and the 95 and 99 percentiles of the results of each tumour marker were calculated.

Sensitivity of CEA, CA 19-9 and CA 72-4

In gastric cancer, colorectal cancer and pancreatic cancer the sensitivity was calculated at the 95th and 99th percentile of healthy subjects. The same calculations were performed using the reference values based on benign gastrointestinal disease. Significance was calculated using *Fisher*'s exact test.

Results

Intra-assay precision

Every institute performed the determinations on patient samples containing different levels of markers. A target coefficient of variation of less than 5% was reached in 80% of all serum samples including all tumour markers (111 out of 139 = 80%). In serum samples starting at the cut-off level or higher, a coefficient of variation of less than 5% was confirmed in 72 out of 76 samples (95%). All tumour markers showed the same pattern. Examples are given in figures 1 and 2, for CA 19-9 and Cyfra 21-1, respectively.

Inter-assay precision

The target coefficient of variation of less than 10% was confirmed in 86% of all series (N = 120 for the total of the two tumour marker control levels). The inter-assay precision of the tumour marker control levels were determined according to the evaluation protocol. In tables 1 and 2 the range of the inter-assay coefficient of variation of the different markers are represented together with the number of participating laboratories. Also the mean of the coefficient of variation in the different laboratories was calculated. This was between 5.9% and 9.6% using the low control and between 3.9% and 7.2% using high control. Also the number of laboratories with a coefficient of variation of less than 10% is indicated. The lowest mean CV% was found for AFP and the highest for CA 72-4 in both controls.

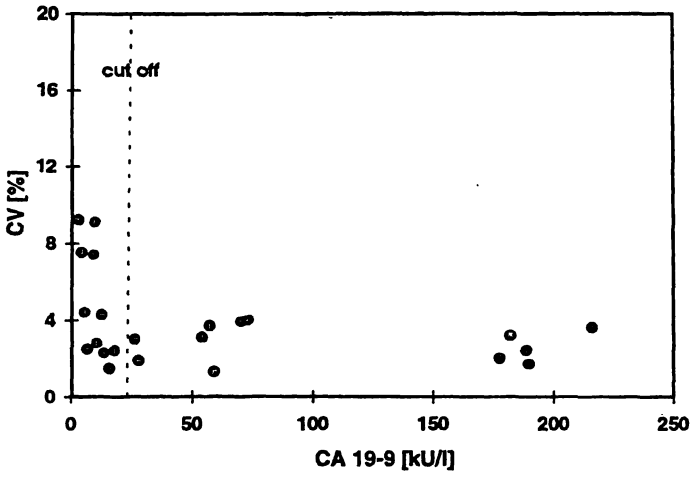

Fig. 1 Enzymun-Test® CA 19-9. Intra-assay precision in human sera as obtained in different institutes. The CV% of each run (y-axis) is plotted against the mean concentration of the sample, N = 23.

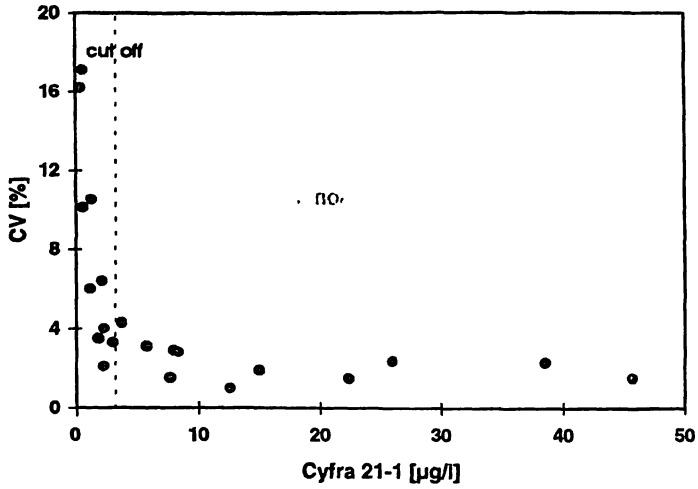

Fig. 2 Enzymun-Test® Cyfra 21-1. Intra-assay precision in human sera as obtained in different institutes. The CV% of each run (y-axis) is plotted against the mean concentration of the sample, N = 21.

Interlaboratory survey

The results for the assayed tumour markers in samples A and B, which were provided by the German Society for Clinical Chemistry (8), were compared with the results obtained in the 'German Ringversuch 1992' using the Boehringer Mannheim Assays. The results are summarized in tables 3a and b. An example is given for CA 19-9 in figure 3. The coefficient of variation was around 13% in the present study (13 laboratories) as compared with around 17% in the 'German Ringversuch 1993' (47 laboratories).

Method comparison studies

In 29 cases results obtained with the ES 300 or the ES 600/700 were compared with those from one of the other methods as performed routinely in the institute (see the section Assays). As examples, the results are presented for CA 15-3 (using the same antibodies) and for CEA (using different antibodies). In three laboratories a total of 283 serum samples were analysed with the CA 15-3 assay of Boehringer Mannheim and CIS bio international. The resulting regression equation was: $y = 1.180\,x - 2.620$ (BM = y and CIS = x). The correlation coef-

Tab. 1 Tumour Marker Control low: The number of laboratories (N) in which the inter-assay precision CVs (with min. and max. ranges) were determined. The number of laboratories with a CV of less than 10% is given. The target values are taken from the package inserts.

Enzymun-Test®		N	Target value	CV [%] $\bar{x}$	CV < 10% N	CV [%] range
CEA	µg/l	13	3.74	6.5	10	1.0–11.6
CA 15-3	kU/l	7	21.7	6.2	6	3.5–10.4
CA 19-9	kU/l	9	22.4	7.9	8	1.8–25.5
CA 72-4	kU/l	8	7.1	9.6	5	1.9–17.6
CA 125 II	kU/l	8	33.9	6.7	7	0.5–14.2
Cyfra 21.1	µg/l	11	4.9	7.6	9	2.3–24.3
AFP	kU/l	4	7.3	5.9	3	2.1–14.5

Tab. 2 Tumour Marker Control high: The number of laboratories (N) in which the inter-assay precision CVs (with min. and max. ranges) were determined. The number of laboratories with a CV of less than 10% is given. The target values are taken from the package insert.

Enzymun-Test®		N	Target value	CV [%] $\bar{x}$	CV < 10% N	CV [%] range
CEA	µg/l	13	38.6	4.9	13	0.8– 8.7
CA 15-3	kU/l	7	74.9	5.4	7	1.5– 9.9
CA 19-9	kU/l	9	84.0	5.1	8	2.5–10.4
CA 72.4	kU/l	8	44.4	7.2	6	2.4–12.0
CA 125 II	kU/l	8	93.1	6.6	7	3.7–11.3
Cyfra 21.1	µg/l	11	34.8	5.3	10	1.7–11.9
AFP	kU/l	4	77.1	3.9	4	0.7– 6.9

ficient (r) was 0.967 ($p = < 10^{-6}$). Figure 4 shows the plotted results from 263 serum samples within the measuring range of the Boehringer Mannheim assay. In four laboratories 213 serum samles were analysed for CEA with the Abbott IMx assay and the Boehringer Mannheim. The correlation is represented in figure 5. The regression equation was: $y = 1.136\ x + 0.523$ (BM = y and IMx = x) ($r = 0.967$; $p = < 10^{-6}$).

Tab. 3a Interlaboratory survey, TM 2/92 sample A (a), B (b). The median and scatter given as the 16th and 84th percentiles of the results from this study in comparison with the results obtained in the "Ringversuch 1992".

Enzymun-Test®		16%	Median	84%	N
CEA	TM 2/92	13.3	14.4	15.0	66
	Study	13.0	14.5	15.3	25
CA 15-3	TM 2/92	14.0	15.9	17.8	51
	Study		17.3		7
CA 19-9	TM 2/92	10.7	14.2	16.1	47
	Study	12.8	14.6	17.2	13
CA 125	TM 2/92[a]	18.9	22.9	30.5	31
	Study[b]	12.6	15.0	17.9	12
AFP	TM 2/92	7.6	8.8	9.8	27
	Study		9.5		9

[a] 1st generation
[b] 2nd generation

Tab. 3b

Enzymun-Test®		16%	Median	84%	N
CEA	TM 2/92	5.9	6.5	6.7	66
	Study	6.0	6.5	7.1	25
CA 15-3	TM 2/92	11.8	13.4	15.4	51
	Study		15.5		7
CA 19-9	TM 2/92	11.8	15.1	16.8	47
	Study	13.7	15.2	18.4	13
CA 125	TM 2/92[a]	18.6	22.5	26.6	31
	Study[b]	13.9	14.9	16.7	12
AFP	TM 2/92	13.5	15.2	16.3	27
	Study		16.9		9

[a] 1st generation
[b] 2nd generation

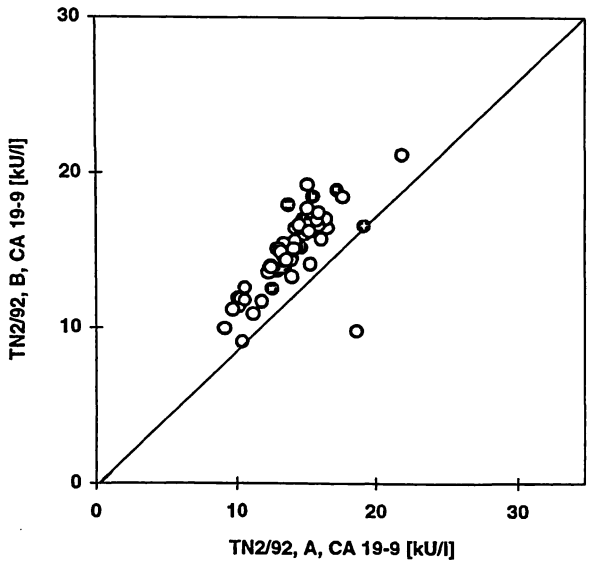

Fig. 3 Interlaboratory survey Enzymun-Test® CA 19-9. A *Youden* diagram with identification of the values obtained in this study ● (13 laboratories) and the values from the Ringversuch 1992 o with Enzymun-Test® CA 19-9.

Reference values

Reference values based on helathy subjects were determined using the Boehringer Mannheim assay. The results were compared with the value on the package insert of the assays. Table 4 shows that all results obtained

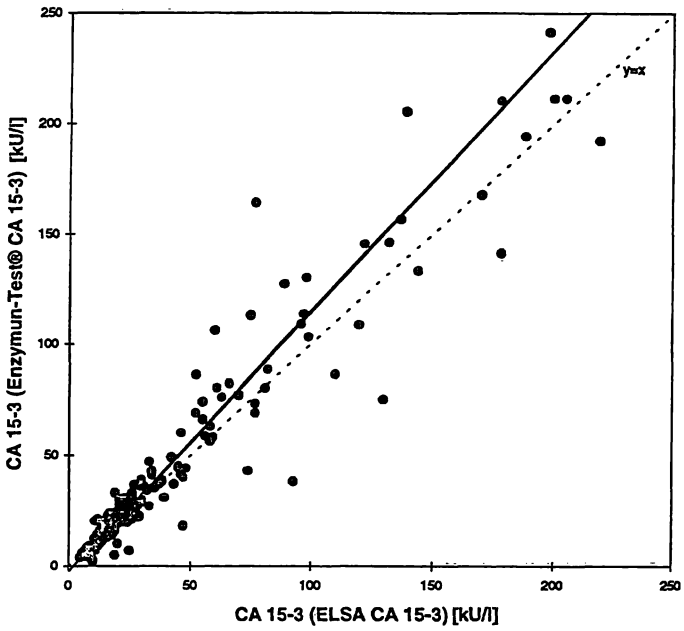

Fig. 4 CA 15-3. Comparison of Enzymun-Test® CA 15-3 (y-axis) with ELSA CA 15-3 (x-axis). The correlation is indicated by a solid line in comparison with y = x.

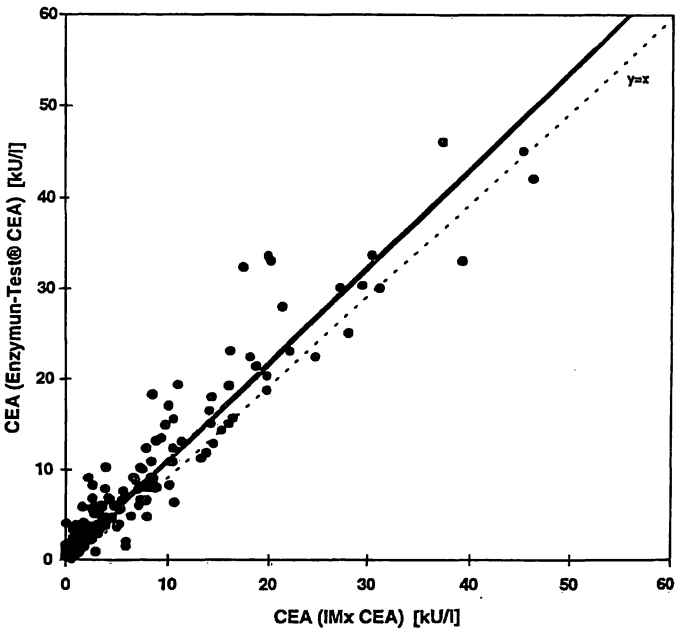

Fig. 5 CEA. Comparison of Enzymun-Test® CEA (y-axis) with IMx CEA (x-axis). The correlation is indicated by a solid line in comparison with y = x.

Tab. 4 Reference values for 7 tumour markers. 95th and 99th percentiles of previous experiments and results of the study determined in healthy subjects.

Quantity Enzymun-Test®		Package insert		N	Results of the study	
		95th percentile	99th percentile		95th percentile	99th percentile
CEA	µg/l	4.6	5.7	351	3.4	4.4
CA 15-3	kU/l	22	30	44	21	22
CA 19-9	kU/l	22	37.0	166	21.5	34
CA 72-4	kU/l	6.7	9.8	135	4.5	9.3
CA 125 II	kU/l	35	65	104	22	35
Cyfra 21.1	µg/l	3.3[1])		135	8.8[1])	
		1.8[2])		208	2.2[2])	

[1]) values based on non-malignant lung diseases

[2]) values based on healthy subjects

in this study (with the exception of Cyfra 21-1) were below the assigned values. The reference value of Cyfra 21-1 based on non-malignant lung diseases (8.8 µg/l) is significantly higher than the value quoted on the package insert.

CA 125 in gynaecological diseases

The 95th percentile in healthy female blood donors (N = 140) using Enzymun-Test® CA 125 II was 21.8 kU/l. This is substantially lower than the generally accepted reference level of 35 kU/l. Of 41 patients with endometriosis, 28 (68%) had levels above 21.8 kU/l and 25 patients (61%) had concentrations above 35 kU/l. Levels of CA 125 II up to 295 kU/l were recorded in patients with endometriosis. In contrast, in patients with benign ovarian cysts (N = 20), 7 (35%) had concentrations above 21.8 kU/l and only 2 (10%) had levels above 35 kU/l.

Sensitivity of tumour markers in gastrointestinal cancer

In table 5 the 95% reference levels of CEA, CA 19-9 and CA 72-4 are presented as they were determined in this study in healthy individuals (N = 1098) and benign gastrointestinal diseases (N = 103). The sensitivity of these markers at 95% specificity, as determined in this study for normals and patients with benign disease, was calculated in pre-operative samples of patients with colon cancer, pancreatic cancer and gastric cancer. These results are summarized in Tables 6a, b, and c.

Tab. 5 Reference values of Enzymun-Test® CEA, CA 19-9 and CA 72-4 based on healthy subjects obtained in this study and values related to patients with benign gastrointestinal diseases.

	Enzymun-Test®		
	CEA [µg/l]	CA 19-9 [kU/l]	CA 72-4 [kU/l]
Healthy subjects	3.4	21.5	4.5
Benign gastrointestinal diseases, N = 103	6.2	75	8.9

In colon cancer, CEA is the most sensitive marker for both reference groups: 59% and 44% respectively. The difference in sensitivity with either CA 19-9 or CA 72-4 is highly significant ($p = 0.008$ and $p = 5.2 \times 10^{-7}$, respectively). The combination of CEA with CA 19-9 increases the sensitivity to 70% using healthy individuals as the reference group ($p = 0.02$). In comparison with the benign gastrointestinal disease group a small increase is seen by combining CEA with CA 19-9 or CA 72-4. However, this is not significant. In pancreatic cancer, CA 19-9 showed the highest sensitivity, compared with both reference groups (78% and 62%, respec-

Tab. 6 Gastrointestinal cancers. The sensitivity was calculated according to the cut-off based on healthy subjects obtained in this study and the cut-off based on patients with benign gastrointestinal diseases. The number expresses the number of values above the cut-off; in ( ) the percentage elevated values is given.

Property	Normal cut off		Benign cut off	
a) Colon cancer, N = 226				
CEA	3.4 µg/l	134 (59)	6.2 µg/l	100 (44)
CA 19-9	21.5 kU/l	105 (47)	75 kU/l	50 (22)
CA 72-4	4.5 kU/l	80 (35)	8.9 kU/l	54 (24)
CA 19-9 a/or CEA		159 (70)		106 (47)
CA 72-4 a/or CEA		132 (58)		110 (49)
CA 19-9 a/or CA 72-4		131 (58)		73 (32)
b) Pancreas cancer, N = 66				
CEA	3.4 µg/l	31 (49)	6.2 µg/l	13 (21)
CA 19-9	21.5 kU/l	49 (78)	75 kU/l	39 (62)
CA 72-4	4.5 kU/l	21 (33)	8.9 kU/l	15 (24)
CA 19-9 a/or CEA		55 (87)		43 (68)
CA 72-4 a/or CEA		39 (62)		23 (37)
CA 19-9 a/or CA 72-4		50 (79)		41 (65)
c) Gastric cancer, N = 110				
CEA	3.4 µg/l	46 (42)	6.2 µg/l	29 (26)
CA 19-9	21.5 kU/l	48 (44)	75 kU/l	27 (25)
CA 72-4	4.5 kU/l	40 (36)	8.9 kU/l	34 (31)
CA 19-9 a/or CEA		63 (57)		39 (36)
CA 72-4 a/or CEA		54 (49)		43 (39)
CA 19-9 a/or CA 72-4		62 (56)		40 (36)

tively with p-values ranging from 7.5×10^{-4} to 4.3×10^{-7}). The small increase in sensitivity achieved by combining CA 19-9 with either CEA or CA 72-4 is not significant.

In gastric cancer, CA 19-9 showed the highest sensitivity (44%) compared with healthy individuals, whereas CA 72-4 has the highest sensitivity (31%) compared with the benign disease group. However, these differences in sensitivity between the three markers are not significant. With healthy individuals as the control group, the sensitivity is increased by combining CA 19-9 with either CEA or CA 72-4 ($p = 0.03$ and $p = 0.04$, respectively). However, there is no significant difference between the combination CA 19-9 + CEA or CA 72-4 + CEA. With the benign disease group as control, sensitivity is highest in the combination CA 72-4 + CEA but this is not significantly different from the sensitivity of the CA 72-4 determination alone. In 74 out of 110 serum samples from gastric cancer patients, staging was available (tab. 7). The sensitivity of all markers in stage I and II cancer is very low, compared with the healthy control group. When compared with the benign disease group, only CA 19-9 showed elevated levels (6/19) in stage II cancer. In stage III cancer the sensitivity of the three markers is the same for both reference groups. In stage IV, CA 72-4 has the highest sensitivity (59%) as compared with the benign disease group. This is significant when compared with CA 19-9 ($p = 0.04$), but not when compared with CEA. When the benign disease group is taken as the reference group, only CA 72-4 is stage-dependent (stage III/IV: $p = 0.04$).

Discussion

Quality control and standardisation of tumour marker tests is becoming increasingly important, since it is important, for both patients and physicians, to use the results of these marker tests appropriately, and to avoid uncritical overestimation or premature rejection of their use (9–11). The trend toward increasing automation is favorable with respect to reproducibility of test results. Also our results from this multicentre study, using the automatic ES 300 and ES 600 analysers, prove that within a given method highly satisfactory results can be obtained with tumour marker determinations. In 95% of cases, the target intra-assay coefficient of variation of less than 5% was achieved for all markers in serum samples with levels starting at the cut-off level and higher. In 86% of the inter-assay determinations in the institutes using low and high control samples, a target inter-assay coefficient of variation of less than 10% was found (tabs. 1 and 2). Also the results of the interlaboratory survey demonstrate the consistency of the test results of the system used (tabs. 3 and 4; fig. 3). Linearity of the test system was already reported to be highly satisfactory (5).

However, it is obvious from figures 4 and 5 that the determination of the same tumour marker using different test systems is not well standardised, even when the same monoclonal antibodies are used, as in the CA 15-3 tests (12). The problem of standardisation increases when different antibodies are used, as in the CEA test.

Of special clinical interest are the reference levels determined in different institutes compared with the values given by the manufacturer in the package insert. In this study, combining the tumour marker values of healthy individuals as they were determined in 23 institutes, the 95 and 99 percentiles was somewhat lower than given by the manufacturer of Enzymun-Test®. This is again an indication that the system is highly reproducible in different institutes. Only the reference limit of Cyfra 21.1 was found to be somewhat higher in healthy individuals, and this reference limit was definitely higher when patients with benign lung disesae were taken as the reference group (in accordance with the manufacturer's insert). To find the correct cut-off in benign lung diseases it will be necessary to use larger groups, including a clear definition of diseases like chronic obstructive lung disease, acute obstructive lung disease, inflammatory lung disease etc. Therefore the sensitivity of CYFRA 21.1 in lung cancer needs further investigation (13).

CA 125 discriminated well between patients with endometriosis and healthy female donors (14). In our study using the second generation test of CA 125, the 95th percentile of healthy female donors was distinctly lower (21.8 kU/l) than reported for the original Centocor CA 125 version (35 kU/l).

CEA has long been used in colorectal cancer, and more recently the determination of CA 19-9 was applied in pancreatic cancer (15–17). The contribution of the new

Tab. 7 The number of patients with elevated levels of CEA, CA 19-9 and CA 72-4 in gastric cancer stage I–IV according to the reference is represented. Values based on healthy subjects obtained in this study and on patients with benign gastrointestinal diseases. The percentage of elevated levels is given between ( ).

Gastric cancer		N	Enzymun-Test®		
Stages			CEA [µg/l]	CA 19-9 [kU/l]	CA 72-4 [kU/l]
I	normal	6	0	0	1
	benign		0	0	0
II	normal	19	6 (32%)	6 (32%)	1 (0.5%)
	benign		0	6 (32%)	0
III	normal	17	8 (47%)	7 (41%)	7 (41%)
	benign		5 (29%)	4 (24%)	5 (29%)
IV	normal	32	19 (59%)	19 (59%)	20 (63%)
	benign		15 (47%)	11 (34%)	19 (59%)

marker CA 72-4 has been discussed, and its application in gastric cancer has been proposed (18–26). We compared the results of these three markers in gastrointestinal cancers in comparison with the cut-off in healthy individuals and in benign gastrointestinal diseases. As expected, a higher cut-off was determined for the three markers using as the reference group patients with benign gastrointestinal diseases (tab. 5). In particular, CA 19-9 showed a much higher cut-off (75 kU/l versus 21.5 kU/l) for this reference group. In colorectal cancer (including 226 patients with all preoperative stages) CEA appeared to be still the marker of choice, irrespective of the reference group. By combining marker results – where one or the other should be elevated – the specificity was decreased. The increase in sensitivity was only significant when the combination of CEA and CA 19-9 was used, as compared with the normal cut-off. This is caused by the low specificity of CA 19-9 in benign gastrointestinal disease. In pancreatic cancer (63 patients) CA 19-9 was the most sensitive marker, as already reported (15–17). No significant increase in sensitivity is obtained by combining CA 19-9 and CEA determinations or CA 19-9 and CA 72-4. It has been reported that CA 72-4 has a high sensitivity in gastric cancer (18, 19). Others, in rather small series, could not confirm this observation, but they used different cut-off levels, i. e. 4 kU/l (20) and 6 kU/l (21). In a larger series, *Wobbes* et al. used a cut-off of 3 kU/l and estimated a sensitivity for both CA 72-4 and CA 19-9 (cut-off 37 kU/l) of 34% in a group of 94 patients with gastric carcinoma (22). In our study, comprising 110 patients, we determined the sensitivity of CA 72-4, CA 19-9 and CEA in comparison with the 95% reference level in normals and in benign gastrointestinal disease from our own investigations. For CEA these reference levels were somewhat lower than is generally accepted (3.4 µg/l in normals and 6.2 µg/l in benign disease). For CA 19-9 the reference level in normals was lower (21.5 kU/l) but in benign disease we found a high reference level of 75 kU/l. For CA 72-4 the reference level in normals was 4.5 kU/l and in benign disease 8.9 kU/l. Therefore the sensitivity of these markers does not compare very well with that reported earlier (20–22). However, the problem of standardisation of test results using different test systems has to be kept in mind. Most of the investigations with CA 72-4 and CA 19-9 were performed using the Centocor IRMA test. In comparison with healthy individuals CA 19-9 showed the highest sensitivity (44%), but compared with the benign disease group CA 72-4 was more sensitive (31%). This can be explained by the high reference level of CA 19-9 in the benign disease group. In our calculations the differences in sensitivity were not significant. However, significance was not calculated in the earlier reports (20–22). Besides, the cut-off levels chosen in these studies for the different markers resulted in different specificities of the markers. In a study including 161 gastric patients, divided according to stage, *Guadagni* et al. found CA 72-4 to be more sensitive (42.2%) than CA 19-9 (32.3) or CEA (24.2%) in the overall group (23). The highest sensitivity was reported in stage IV (including metastatic disease): CA 72-4: 58%; CA 19-9: 44.2% and CEA: 39.5%. This is in agreement with our results where a sensitivity for CA 72-4 of 59% was determined in stage IV, which was higher than the sensitivity of CA 19-9 (34%) using the benign disease group as the control. They also stated that the sensitivity of CA 72-4 was stage-dependent, as we also found in our study. We also confirmed that the sensitivity could be increased by combining CA 72-4 and CA 19-9 (in our study only when the healthy controls were taken as the reference group). Other recent data in large patient groups indicated CA 72-4 to be the marker of choice using a cut-off level for CA 72-4 of 3.9 kU/l at 95% specificity in a group of patients with benign disease of the gastrointestinal tract, including patients with benign liver disease (24). However, at this 95% specificity the cut-off level for CA 19-9 was very high: 166 kU/l. This could explain the relatively high sensitivity of CA 72-4 (36%) compared with CA 19-9 (21%). At the time of local relapse, or occurrence of distant metastases, the sensitivity of CA 72-4 was 56% (56%) and of CA 19-9 18% (28%). *Gartner* et al. could not confirm this high sensitivity of CA 72-4 compared with CA 19-9 (25). But in this study the cut-off value for CA 72-4 was 6.8 kU/l and for CA 19-9 37 kU/l. *Safi* et al. calculated again a high sensitivity for CA 72-4 using different cut-off levels for CA 72-4 i. e. 2.5 kU/l and 10 kU/l (26). Very few patients with benign gastrointestinal disease showed levels above 10 kU/l. Probably the reference level of CA 19-9 was fixed at 37 kU/l. They also found CA 72-4 to be clearly stage-dependent, but determined a sensitivity for CA 72-4 of 31% in stage I gastric cancer versus 88% in stage IV.

In conclusion, in the present study we compared the sensitivity of the markers CEA, CA 19-9 and CA 72-4 in gastrointestinal cancer at 95% specificity, using cut-off levels of healthy individuals and patients with benign gastrointestinal disease, according to earlier recommendations by the working group 'Quality control and Standardization of Tumour Marker Tests' under the auspices of the Hamburg Symposia on Tumour Markers (27). The reference data were obtained in the study and not just taken from the test package inserts. CEA is still the marker of choice in colorectal cancer as is CA 19-9 in pancreatic cancer. In these two types of cancer no increase in sensitivity was registered by combining markers when the benign disease group was taken as the reference group. Therefore it seems sufficient to determine only one marker in these disease groups. In gastric cancer no definite conclusion can be given. The advantage of CA 72-4 could be that the marker is stage-depen-

dent and has the highest sensitivity in advanced disease (stage IV) as compared with the benign disease group.

The intra-assay precision, inter-assay precision and interlaboratory survey all proved that Enzymun-Test® gave reliable results for tumour marker tests. From the analytical point of view, this fulfills a further requirement of the previously mentioned working group (27). Recently, at the *Bergmeier* conference under the auspices of IFCC, these requirements were discussed for several tumour markers (9, 12). We believe that in the near future unequivocal analysis and reporting of tumour marker data will be a necessity in order to supply clinicians with relevant information for the care of oncological patients.

References

1. Gosling JP. A decade of development in immunoassay methodology. Clin Chem 1990; 36/8:1408–27.
2. Geilenkeuser WJ, Kruse R, Röhle G. Ringversuche der Deutschen Gesellschaft für Klinische Chemie – Interpretation der Auswertung. DG Klinische Chemie Mitteilungen 1992; 23:57–81.
3. Dahlmann N. Qualitätssicherung bei der Bestimmung von Tumormarkern: Darstellung nach Diskussion eines Ringversuches. Lab Med 1991; 15:344–6.
4. Zucchelli GC, Pilo A, Cohen R, Cianetti A, Torre GC. Tumor markers: quality control and standardization in Italy. In: Klapdor R, editor. Current tumor diagnosis: applications, clinical relevance, research, trends. München: Zuckschwerdt, 1994:488–91.
5. Zucchelli GC, Pilo A, Jaworek D, Masini S, Chiesa MR. Immunoassay of CEA, CA 19-9, CA 125 and CA 15-3 on the automated systems ES 300 and ES 600: methodological evaluation from a multicentre collaborative study. Eur J Clin Chem Clin Biochem 1992; 30:875–9.
6. Fiedler H, Kessler A-Ch, Köhler P. Results of a multicentre study on the analytical reliability of the tumour markers CEA, CA 15-3, CA 19-9, CA 72-4 and CA 125. Klin Lab 1994; 40:873–80.
7. Passing H, Bablok WL. A new biometrical procedure for testing the equality of measurements from two different analytical methods. J Clin Chem Clin Biochem 1983; 21:709–20.
8. Kruse R, Geilenkeuser W-J, Röhle G. Interlaboratory surveys of the determination of tumour markers scatter and repeatability of the results. Eur J Clin Chem Clin Biochem 1993; 31:139–46.
9. Schwartz MK. Current status of tumour markers. Scand J Clin Lab Invest 1995; 55 Suppl 221:5–14.
10. Bodenmüller H. The biochemistry of CYFRA 21-1 and other cytokeratin-tests. Scand J Clin Lab Invest 1955; 55 Suppl 221:60–6.
11. Jäger W, Kramer S, Palapelas V, Norbert L. Breast cancer and clinical utility of CA 15-3 and CEA. Scand J Clin Lab Invest 1995; 55 Suppl 221:87–92.
12. Dalen A van. Analytical requirements and standardization of CA 15-3. Scand J Clin Lab Invest 1995; 55 Suppl 221:102–4.
13. Stieber P, Bodenmüller H, Banauch D, Hasholzer U, Dessauer A, Ofenloch-Hähnle B, et al. Cytokeratin 19 fragments: a new marker for non-small-cell lung cancer. Clin Biochem 1993; 26:301–4.
14. Barbieri RL, Niloff JM; Bast Jr RC, Schaetzl E, Kistner RW, Knapp RC. Elevated serum concentrations of CA 125 in patients with advanced endometriosis. Fertility and Sterility 1986; 45:630–4.
15. Bac DJ, Kok TC, Gaast A van der, Splinter TA. Evaluation of CA 19-9 serum levels for monitoring disease activity during chemotherapy of pancreatic carcinoma. J Cancer Res Clin Oncol 1991; 117:263–5.
16. Lambiase L, Vogel SB, Forsmark CE. Elevated CA 19-9 predicts unresectable pancreatic cancer [abstract]. Gastroenterology 1993; 104 (Suppl 4):418.
17. Pleskow DK, Berger HJ, Gyves J, Allen E, McLean A, Podolsky DK. Evaluation of a serologic marker, CA 19-9, in the diagnosis of pancreatic cancer. Ann Int Med 1989; 110:704–9.
18. Klug TL, Sattler MA, Colcher D, Schlom J. Monoclonal antibody immunoradiometric assay for an antigenic determinant (TAG 72) on a novel pancarcinoma antigen (TAG-72). Int J Cancer 1986; 38:661–9.
19. Byrne DJ, Browning MCK, Cuscieri A. CA 72-4: a new marker for gastric cancer. Br J Surg 1990; 77:1010–3.
20. Kornek GV, Depisch D, Rosen HR, Temsch EM, Scheithauer W. Comparative analysis of CA 72-4, CA 195 and carcinoembryonic antigen in patients with gastrointestinal malignancies. J Cancer Res Clin Oncol 1992; 118:318–20.
21. Yasasever V, Sengün Z, Saydan N, Onat H, Dalay N. Serum values of CA 72-4 in patients with gastrointestinal system tumors comparison with CEA and CA 19-9. Eur J Gynaec Oncol 1992; XIII (5):403–8.
22. Wobbes T, Thomas CMG, Segers MFG, Nagengast FM. Cancer 1992; 69(8):2036–41.
23. Guadagni F, Roselli M, Cosimelli M, Ferroni P, Spila A, Casaldi V, et al. Correlation between positive CA 72-4 serum levels and lymph node involvement in patients with gastric carcinoma. Anticancer Res 1993; 13:2409–14.
24. Reiter W, Stieber P, Reuter C, Cramer C, Fateh-Moghadam A. CA 72-4, CEA, CA 19-9 in diagnosis and relapse and metastases and in prognosis of stomach cancer patients. In: Klapdor R, editor. Current tumor diagnosis: applications, clinical relevance, research, trends. München: Zuckschwerdt Verlag, 1994:13–6.
25. Gärtner U, Scheulen ME, Aghabi E, Wiefelspütz J, Delbrück H. Determination of CA 72-4 by IRMA or ELISA in the follow-up of gastric cancer patients. In: Klapdor R, editor. Current tumor diagnosis: applications, clinical relevance, research, trends. München: Zuckschwerdt Verlag, 1994:17–9.
26. Safi F, Kuhns V, Beger HG. Tumor markers in gastric cancer. In: Klapdor R, editor. Current tumor diagnosis: applications, clinical relevance, research, trends. München: Zuckschwerdt Verlag, 1994:20–7.
27. Dalen A van. Quality control and standardization of tumor marker tests. Tumor Biol 1993; 14:131–5.

Received June 21, 1995/January 7, 1996

Corresponding author: Dr. A. van Dalen, Groene Hart Ziekenhuis, Bleuland location, Department of Nuclear Medicine, P. O. Box 1098, NL-2800 BB Gouda, The Netherlands

Eur J Clin Chem Clin Biochem 1996; 34:385–386

Report on the Symposium[1])
"Drug Effects in Clinical Chemistry Methods"

held on December 8, 1995 in Penzberg, Bavaria, Germany

Josef Breuer

Marienhospital Gelsenkirchen, Gelsenkirchen, Germany

Summary: The aim of the symposium was to establish a list of 20–30 drugs and to determine test concentrations (at therapeutic levels and above) that would indicate interference to clinical chemistry methods in serum and plasma. The following agents were chosen:

Acetaminophen, Acetylcysteine, Acetylsalicylic acid, Ampicillin, Ascorbic acid, Ca-Dobesilate, Cefoxitin, Cyclosporine, Heparin, Ibuprofen, Intralipid, Levodopa, Methyldopa, Metronidazole, Phenylbutazone, Rifampicin, Tetracycline, Theophylline.

Introduction

On December 8, 1995, the following experts came together in Penzberg, Bavaria, Germany, to discuss drug effects on clinical chemistry methods:

J. Breuer, Gelsenkirchen, Germany
R. Galimany, Barcelona, Spain
P. Gerthoux, Milano, Italy
P. Koller, Mannheim, Germany
W. Mühe, Gelsenkirchen, Germany
J. Salway, Surrey, Great Britain
R. Scholer, Basel, Switzerland
O. Sonntag, Illkirch, France
N. Tryding, Kristianstad, Sweden

The aim of this symposium was to establish a drug interference list and testing levels at which one may determine interference to clinical chemistry methods in se-

Tab. 1 Drugs and concentrations to be tested for drug interferences in serum/plasma

Active compound	Clinical Use	Test concentration[a] c_1 (mg/l)	Test concentration[b] c_2 (mg/l)
Acetylcysteine	Mucolytic	150	30
Ampicillin	Antibiotic	1000	200
Ascorbic acid	Vitamin	300	30
Ca-Dobesilate	Vasotherapeutic	200	20
Cefoxitin	Antibiotic	2500	250
Heparin	Anticoagulant	500 IU/l	50 IU/l
Intralipid	Supplement	10000	2000
Levodopa	Antiparkinson	20	4
Methyldopa	Antihypertonic	20	2
Metronidazole	Chemotherapeutic	200	10
Phenylbutazone	Analgetic	400	100
Rifampicin	Chemotherapeutic	60	20
Acetylsalicylic acid	Analgesic	1000	300
Acetaminophen	Analgesic	200	20
Cyclosporine	Immunosuppressive	5	1
Ibuprofen	Analgesic	500	50
Tetracycline	Chemotherapeutic	50	10
Theophylline	Coronary vasodilator	100	10

[a] c_1 drug concentration above therapeutic level
[b] c_2 drug concentration at therapeutic level

[1]) The symposium was generously supported by Boehringer Mannheim.
A further symposium is scheduled to make a list of drugs for in vitro testing in urine.

rum and plasma and which can be tested in vitro. To prepare for the discussion, everyone had to send a list of his "favorite" drugs to the organizers. Additional lists were also sent by *G. Siest*, Nancy, France and *H. Sine*, Indianapolis, USA.

Results and Discussion

At the beginning of the meeting the following criteria were examined in order to select a valid list from the above mentioned "favorite" lists:

- Typical (high) serum concentration
- Typical test interference on clinical chemistry methods
- How often is the drug used?
- Relevance in vivo
- Drugs absorbing light at clinical chemistry methods wavelengths

Further background information utilized during the discussion was the 7th edition of the book by *Nils Tryding* (1) and the proposed Guideline Interference Testing in clinical chemistry by NCCLS (2).

The outcome of the symposium is summarized in table 1.

References

1. Tryding N, Tufvession C, Sonntag O. Drug effects in clinical chemistry, 7th ed. Apoteksbolaget, Stockholm, 1996.
2. National Committee for Clinical Laboratory Standards Interferences Testing in Clinical Chemistry. NCCLS Socument EP7-P, Vol. 6 No 13. ISSN 0273-3099, 1986.

Received January 23, 1996

Corresponding author: Prof. Dr. J. Breuer, Zentrallaboratorium, Marienhospital, Virchowstraße 135, D-45886 Gelsenkirchen, Germany