Computers and Biomedical Research
Volume 25, Number 6, December 1992

CONTENTS

Vol. 25, No. 6, December 1992 COMPUTERS AND BIOMEDICAL RESEARCH Pages 511–603

ISSN 0010-4809

This Number Completes Volume 25
Volume 25, Number 6, December 1992

Computers and Biomedical Research

AN INTERNATIONAL JOURNAL

An Official Publication of the American Medical Informatics Association
Endorsed by the American College of Medical Informatics

ACADEMIC PRESS, INC.

COMPUTERS AND BIOMEDICAL RESEARCH 25, 511–526 (1992)

Validation of the Medical Expert System PNEUMON-IA[1]

A. VERDAGUER,*,† A. PATAK,* J. J. SANCHO,* C. SIERRA,‡ AND F. SANZ*,[2]

*Departament d'Informàtica Biomèdica, Institut Municipal d'Investigació Mèdica, Facultat de Medicina (Universitat Autònoma de Barcelona), Barcelona, Spain; †Aliança Mataronina, Mataró, Barcelona, Spain; and ‡Centre d'Estudis Avançats de Blanes, Blanes, Barcelona, Spain

Received June 12, 1990

The present study validates the expert system PNEUMON-IA. The aim of PNEUMON-IA is assessing the etiology of community-acquired pneumonias from clinical, radiological, and laboratory data obtained at the onset of the disease. Validation was performed using data from medical records of 76 patients with proven clinical diagnosis of pneumonia. The etiological diagnoses provided by PNEUMON-IA were compared to those established by five specialists unrelated to the development of the expert system. For each etiological possibility, both PNEUMON-IA and the experts provided a causal possibility, expressed as a linguistic label (i.e., "almost impossible"). Linguistic labels were then converted to numeric values. In the majority of cases, an etiological diagnosis was unavailable to be used as a gold standard. To overcome this limitation, distances between arrays of etiological possibilities given by specialists and by PNEUMON-IA were considered as an agreement measure between diagnoses. Cluster analysis based on those distances was used to classify PNEUMON-IA among experts. Results showed the same differences between specialists and PNEUMON-IA as among the specialists themselves. The method used to validate PNEUMON-IA could prove useful to assess the performance of expert systems in fields where no gold standard is available.

INTRODUCTION

The validation of any expert system is a fundamental step in its development (*1*, *2*). To use a diagnostic expert system such as PNEUMON-IA in clinical practice the validity of its diagnoses should be verified. We present a validation of PNEUMON-IA through the etiological diagnoses it provided, without taking into account other characteristics of the system such as its usefulness, user interface, or reasoning explanation. The originality of the validation methodology used lies in the fact that it evaluates an expert system in a field where no absolute gold standard exists as a reference.

[1] This study was supported in part by Grants FIS 87/1387 and 88/1362.

[2] To whom correspondence and reprint requests should be addressed at: Departament d'Informàtica Biomèdica, Institut Municipal d'Investigació Mèdica, Passeig Marítim 25-29, 08003 Barcelona, Spain. Fax: (343) 485 49 52.

0010-4809/92 $5.00

R02013 IF 1) Community-acquired Pneumonia
2) Frequent contacts with animals
3) Bacterial Disease
THEN [QUITE POSSIBLE]
Pneumococcus

R02014 IF 1) Community-acquired Pneumonia
2) Frequent contacts with birds
3) Bacterial Disease
THEN [VERY POSSIBLE]
Pneumococcus

FIG. 1. Examples of two rules of PNEUMON-IA. R02014 is more specific than R02013 thanks to the relation *bird is a kind of animal*.

The Expert System

The aim of PNEUMON-IA is to assess the etiology of community-acquired pneumonias from clinical, radiological, and laboratory data obtained at the onset of the disease (*3–5*). It uses an inference engine named MILORD (*1, 4, 5*) which uses fuzzy logic and linguistic labels to express uncertainty. PNEUMON-IA's knowledge is implemented through production rules (Fig. 1). Rules are grouped in modules and strategies and controlled by metarules in a multilevel architecture represented in Fig. 2. PNEUMON-IA considers the 22 possible etiological agents listed in Table 1. For each case, it provides an output which includes a list of the most likely etiologies, each one qualified by a linguistic label to express its associated possibility (see Appendix).

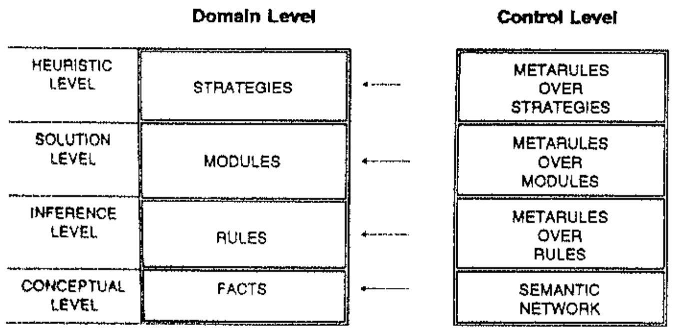

FIG. 2. Multilevel architecture of PNEUMON-IA.

TABLE 1

ETIOLOGIES CONSIDERED BY THE EXPERT SYSTEM PNEUMON-IA

Bacterial pneumonias	Viral pneumonias
Streptococcus pneumoniae	*Respiratory viruses*
Legionella pneumophila	*Chicken pox*
Enterobacteriaceae	*Bacterial superinfection of viral*
Haemophilus influenzae	*Cytomegalovirus*
Pseudomona aeruginosa	
Anaerobic bacteria	Fungal pneumonias
Staphylococcus aureus	*Aspergillus spp.*
Branhamella catarrhalis	*Cryptococcus neoformans*
Streptococcus pyogenes	
Mycobacterium tuberculosis	Pneumonias caused by parasites
Nocardia asteroides	*Pneumocystis carinii*
Mycoplasma pneumoniae	
Chlamydiae spp.	
Q fever pneumonia	
Neisseria meningitidis	

In its present form, PNEUMON-IA knowledge base comprises 487 facts, 659 rules, 92 metarules, and 25 modules. As a programming language we used VAXLISP, a superset of CommonLISP, running on VAX computers under VMS operating system. In addition, there is a CommonLISP version of PNEUMON-IA for SUN workstations running under UNIX operating system.

The Problem

Although of obvious importance for its treatment, etiological diagnoses of pneumonia imply great uncertainty, since etiology is seldom confirmed (Table 2) and therefore it is difficult to establish a gold standard to compare this expert system with. One of the possible methods is based in the selection of a sample including only confirmed cases, as in the validation of MYCIN (*6*). The confirmed cases, however, are often the most severe and are caused by a limited subset of etiological agents. These facts would introduce an unwanted bias in the study.

The Proposed Solution

In the absence of a gold standard, the aim of this validation is to compare, in a sample of patients with community-acquired pneumonia, the etiological approximations proposed by PNEUMON-IA and those proposed by medical specialists. We took the consensus of several medical specialists as a reference, although in a different way to that of other expert systems validation projects

TABLE 2

ETIOLOGIES OF COMMUNITY-ACQUIRED PNEUMONIAS

Reference	n	Pneum.	Legion.	Mycopl.	Virus	Other	UK/UC[a]
(1)	210	11.5	1.5	14.0	15.0	12.5	52.0
(2)	127	76.0	15.0	2.3	8.6	12.5	3.0
(3)	314	22.0	2.5	24.0	21.0	14.0	33.0
(4)	127	54.0	0.7	14.0	7.0	3.3	21.0
(5)	301	18.6	3.9	3.3	16.1	31.6	36.5
(6)	198	39.1	6.2	16.9	4.1	13.5	20.2

Note. Figures in table reflect the percentage of each agent over the total number of agents isolated. The sum of percentages greater than 100 corresponds to polymicrobial etiologies.
[a] UK/UC, unknown/unclear.
[b] Part of the study performed before the identification of *L. pneumophila*.

(*7–10*). Distances between possibility arrays of etiologies given by specialists and PNEUMON-IA were taken as a measure of agreement between them. Cluster analysis of these distances allows us to classify PNEUMON-IA among clinical experts, obtaining a relative assessment of its diagnostic ability.

Although this approach provides results slightly more difficult to interpret, the bias of using only cases with confirmed etiologies does not occur. The agreement with the confirmed diagnosis in the cases that it was known was considered a secondary way of validation.

MATERIALS AND METHODS

1. Sample

It was initially decided to collect a sample of 80 medical records of pneumonia patients from four general hospitals located in the Barcelona area. Four cases were rejected because of incomplete data, leaving 76 cases for analysis. Criteria for inclusion were: clinical and radiological evidence of pneumonia; age equal to or over 14 years; community-acquired pneumonia; and a fully documented medical record. A randomized selection of cases from each hospital's files was performed, stratifying them according to the WHO's ICD code of diagnosis, to ensure that the ratio of different types of pneumonia was preserved.

2. Collection of Data

Collection of data was carried out using a form in which data of possible diagnostic interest and the etiology when confirmed were recorded. Most of the information was related to events occurring during the first 48 hr of the clinical course. Data unavailable in the medical record was classified as unknown. Radiological data were recorded in the form of B&W 35-mm slides for submission to the specialists. The interpretation of the chest X-rays for input to

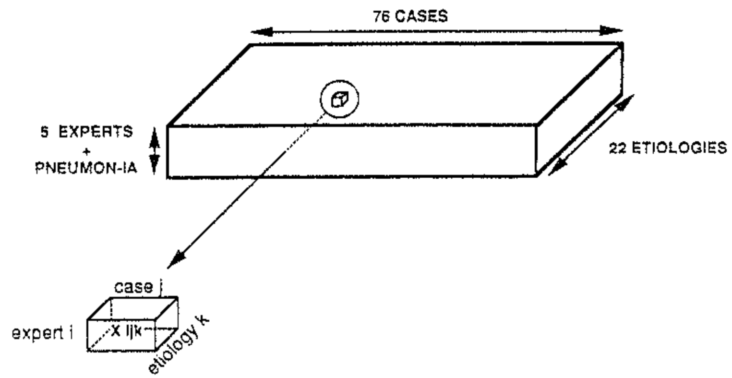

FIG. 3. Cubic matrix, the dimensions of which are the five specialists and PNEUMON-IA, the 76 cases considered, and the 22 etiologies taken into account. x_{ijk} represents one of the eight possible certainty values for a particular etiology attributed to one case by one expert.

the expert system was carried out by the system designer physician because PNEUMON-IA by itself cannot interpret X-rays. Using the intelligent data acquisition module of PNEUMON-IA, data were entered into the computer by one of the authors (A.P.), a physician familiar with PNEUMON-IA operation but unrelated to the development of PNEUMON-IA rules or database. He did not know the treating physician's diagnosis. Data were entered as required by the system if available. The system supplied a set of diagnostics at the end of the process.

3. Assessment of the Cases

PNEUMON-IA gave a diagnosis for each case. Each diagnosis took the form of an array in which the 22 possible etiologies were qualified with one of eight possible labels of possibility. The expert system generated also a validation report for each case which included all data except radiological features and system diagnoses. It was not mandatory for the experts to rate every single possible diagnosis, but it was required for them to give labels of possibility to those considered in the differential diagnosis of each case. PNEUMON-IA provided possibility labels for the diagnostics considered in its search path. These reports together with chest X-ray slides were assessed independently by five specialists from five hospitals. These specialists had not intervened in the development of the expert system. No limit was established on the number of etiologies cited to assess the cases, nor was a limit imposed on the time taken to reach them. A diagram of the data matrix of six experts (PNEUMON-IA + 5 clinical experts), 76 cases, and 22 possible etiologies is shown in Fig. 3.

TABLE 3

Linguistic label	Assigned possibility
Impossible	0.0000
Almost impossible	0.0330
Slightly possible	0.1077
Somewhat possible	0.2416
Possible	0.4500
Quite possible	0.6500
Very possible	0.8486
Sure	1.0000

4. *Concordance Measures between Specialists and the Expert System*

In order to establish the concordance between diagnoses made by specialists and PNEUMON-IA, distances between possibility arrays provided by each specialist and by the expert system were calculated for each case. To compute the distances, linguistic labels were substituted by a number which is a central trend parameter of the range of numeric values compatible with the linguistic labels (*5*) (Table 3).

The following distances were considered (*17*):

— Euclidean

$$d(i,j) = \sqrt{\frac{1}{N}\sum_{m=1}^{N}(x_{im} - x_{jm})^2},$$

where d is the distance, i and j are a pair of experts, N is the total number of etiological possibilities, m is each of the etiological possibilities, and x is the numerically expressed certainty.

— City-block

$$d(i,j) = \frac{1}{N}\sum_{m=1}^{N}|x_{im} - x_{jm}|.$$

City-block distance penalizes the existence of many small differences. In our case city-block distance will magnify the discrepancies between a couple of experts giving slightly different possibilities to a relatively great number of etiological diagnoses.

— Chebychev

$$d(i,j) = \text{Max}_m |x_{im} - x_{jm}|.$$

This distance penalizes the existence of large differences even though they may occur in a low number of etiologies. In our case, Chebychev distance will magnify discrepancies between a couple of experts giving a few disparate etiological possibilities for the same patient.

TABLE 4

MATRIX OF DISSIMILARITIES CALCULATED FROM THE CITY-BLOCK DISTANCE

	E1	E2	E3	E4	E5
E2	0.0495				
E3	0.0636	0.0663			
E4	0.0540	0.0611	0.0700		
E5	0.0581	0.0658	0.0671	0.0702	
ES	0.0566	0.0643	0.0670	0.0659	0.0692

To establish the overall distance between pairs of experts these distances were averaged over the 76 cases considered. Using the obtained 6 by 6 matrix of the averaged distances (Table 4), a hierarchical cluster analysis (*17*) was carried out to condense the information contained in a similarity or dissimilarity matrix by a progressive agglomeration of experts as a function of their similarity. When two experts are clustered, another matrix of dimension $(n - 1) * (n - 1)$ must be built for the process to continue. Every pair of similarity or dissimilarity coefficients of the two clustered experts to the rest of them is then substituted by a single new value. There are several methods for obtaining this new value. In the method presented in this paper, we use the average linkage criteria, which define the new values as an arithmetic mean of the previous ones. Both an advantage and disadvantage of the hierarchical cluster analysis is that the method does not provide a fixed number of clusters but rather a progressive agglomeration of the experts which more or less allows grouping of them by observing the dendrogram closer or further from its vertex. This clustering process has the particularity that all the groups are consistent because the reduction of their number is always generated by aggregating smaller groups. Cluster analysis is a descriptive analysis and does not supply information about the statistical significance of the observed differences between experts, but it is useful enough for the present study because it provides an intuitive relative classification of PNEUMON-IA among human experts. This progressive agglomeration of experts is graphically represented by a "dendrogram," a figure that presents the experts or groups of experts as connected by links placed at a position of the horizontal axis that is proportional to the similarity between experts or groups of experts.

5. *Cross-Evaluation of Specialist's Proficiency*

Specialists were asked to assess their colleagues' proficiency in a confidential report, recording scores for each one on a scale from 1 to 10. Results were averaged and specialists were ranked from highest to lowest proficiency. The scores were: specialist E1 obtained a 10, specialists E2 and E3 obtained scores between 8 and 8.5, and specialists E4 and E5 obtained an average score of 6.5.

TABLE 5

ETIOLOGICAL POSSIBILITIES CONSIDERED IN THE SAMPLE

	E1	E2	E3	E4	E5	ES
S. pneumoniae	62	57	65	65	56	63
L. pneumophila	27	27	30	4	48	27
Enterobacteriaceae	27	27	36	11	25	39
H. influenzae	18	29	20	17	11	16
P. aeruginosa	0	2	0	0	0	0
Anaerobes	22	12	16	15	19	27
S. aureus	7	2	5	8	9	4
B. catarrhalis	0	0	1	0	0	0
S. pyogenes	5	2	1	0	0	0
M. tuberculosis	7	17	6	3	9	3
N. asteroides	1	0	0	1	1	0
M. pneumoniae	10	10	25	10	15	19
Chlamydiae spp.	5	6	19	2	8	29
Q fever	5	7	8	2	7	8
Meningococcus	0	0	0	0	0	0
Viral	1	3	2	1	2	24
Chicken pox	0	0	0	0	0	0
Superinfection	0	0	0	0	0	1
Cytomegalovirus	0	0	0	0	0	0
Aspergillus spp.	1	2	0	0	1	1
C. neoformans	0	0	0	0	0	0
P. carinii	1	1	0	0	0	0
Total	199	204	34	139	215	261
Mean etiology/case	2.62	2.76	3.08	1.83	2.83	3.43

RESULTS

1. *Frequency of Diagnoses*

A descriptive statistical analysis for each specialist was carried out by determining the absolute and relative frequency of the etiologies mentioned with a possibility greater than "almost impossible" (Table 5). Since each specialist usually provided more than one etiological diagnosis for each case, the total number of etiologies recorded was greater than the number of cases. The number of etiologies per case varied between 1.83 for E4 and 3.43 for PNEUMON-IA, the global average being 2.76. The etiology most often cited corresponded to *Streptococcus pneumoniae*, with an average of 61 mentions in 76 cases. The expert system considered pneumonia caused by *Chlamydiae spp.* more commonly than the clinical experts (29 vs an average of 12). However, specialist E3 cited this etiology on 19 occasions. It should be noted that the expert system cited viral pneumonias more often than the specialists did. Although no data were available on the true etiology for the majority of pneumonias, these differences might be due to an underestimation of this etiology by the specialists, since the majority of patients suffering from viral pneumonia are not hospitalized and no effective specific chemotherapy is available.

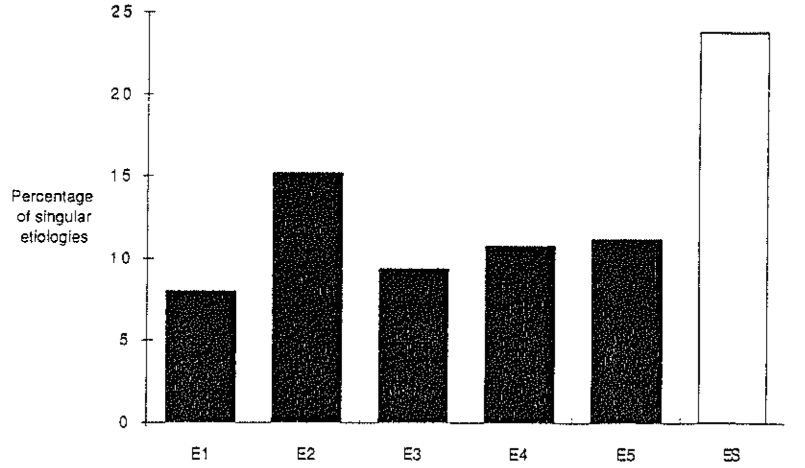

FIG. 4. Percentage of singular etiologies. E1 to E5, human specialists; ES, expert system PNEUMON-IA.

2. *Singular Etiologies*

We considered "singular etiologies" those mentioned only by one specialist in one particular case. One way of determining the concordance of a single specialist with the consensus of the group is to note the occurrence of singular etiologies. Figure 4 shows the percentage of singular etiologies recorded by each specialist. The expert system recorded more singular etiologies (23.8%) than any clinical specialist. However, if viral etiologies are excluded, the percentage is reduced to 15.7%, close to the specialist E2. These results are attributable to the larger number of etiologies provided for each case by PNEUMON-IA, particularly with reference to viral pneumonias as has been pointed out previously.

3. *Omitted Etiologies*

An "omitted etiology" was considered as an etiology mentioned by every specialist (including PNEUMON-IA) but one. The number of omitted etiologies is a measure of disagreement with the consensus. Figure 5 shows the number of etiological diagnoses omitted by each specialist. The number of omitted etiologies by PNEUMON-IA (5 cases) was less than half the omissions made by the specialists such as E2 (12 cases) or E4 (19 cases). This is also an

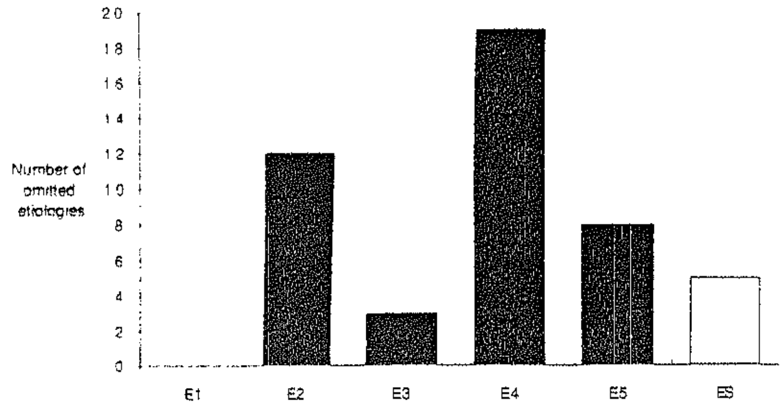

FIG. 5. Number of omitted etiologies. E1 to E5, human specialists; ES, expert system PNEUMON-IA.

indirect consequence of the larger number of etiologies per case proposed by PNEUMON-IA.

4. Confirmed Etiologies

Although only 10 of the 76 cases selected were confirmed by microbiologic examination, it is interesting to analyze the concordance between the etiologies proposed by the specialists and the expert system with this subset of real ones. Figure 6 shows that the confirmed etiology was mentioned by the expert system as the most likely etiologies in 4 occasions (the specialists moved in a range from 3 to 6), and on 7 occasions including also the second most likely diagnosis with a range of 3 to 8 among the group of specialists.

5. Distances between Specialists

5.1. City-block distance. Table 4 shows the 6 by 6 dissimilarity matrix obtained using the city-block distance once all cases were averaged. In Figure 7 we represent graphically as a dendrogram results of a cluster analysis of these distances. As it can be observed in this dendrogram, the distance between specialists E1 and E2 was the smallest. The distance among other specialists progressively increased with no groupings being apparent. PNEUMON-IA occupies a halfway position. Specialists E3 and E5 appeared outside the cluster (E1, E2, E4, PNEUMON-IA) and isolated from each other.

5.2. Euclidean distance. Figure 8 shows the dendrogram of the cluster analysis resulting from a matrix of euclidean distances. Like with city-block distance, specialists E1 and E2 are the closest. The expert system forms a cluster with E1 and E2, other specialists being excluded from this group and isolated from each other.

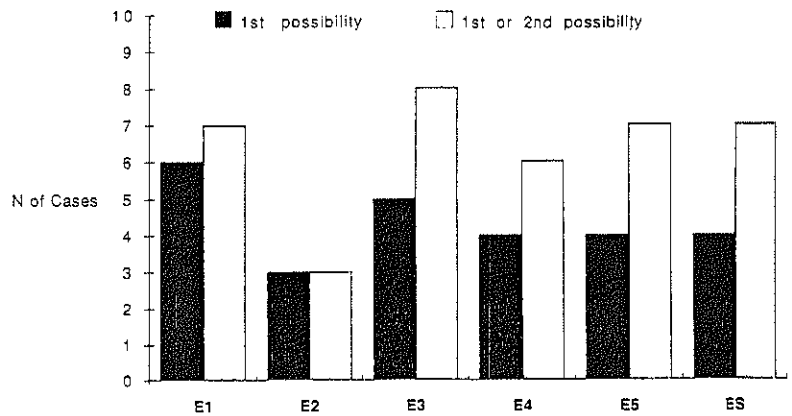

FIG. 6. Number of correct diagnoses in 10 cases with confirmed etiology. E1 to E5, human specialists; ES, expert system PNEUMON-IA.

5.3. Chebychev distance. The dendrogram of the cluster analysis resulting from Chebychev distances was similar to the dendrogram obtained with euclidean distances with regard to specialists E1 and E2 and the expert system. Only the relative positions of specialists E3, E4, and E5 changed.

DISCUSSION

Diagnostic expert systems with output restricted to a single and easily auditable diagnosis are a minority. Their accuracy can be easily compared with that of clinicians (*18, 19*). Well accepted parameters such as sensitivity, specificity, and positive and negative predictive values are often appropriate. Medical expert systems which output a list of diagnoses with their associated possibilities are relatively common. When such systems are applied to a problem where the correct answer is known, the validation process is notably straightforward. Nonetheless, even in the above-mentioned cases, it is astonishing how seldom the computers' and doctors' diagnoses have been directly compared on the same patient on the same occasion using the same indicators (*20*).

Unfortunately, this scenario is unlikely to be customary in modern clinical practice where the correct answer is seldom known for a given problem. Despite this limitation, validation of an expert system requires some measure of success. This means that some approximation to the truth should be known. In the literature, different approximations have been used, such as the opinion of one or more experts, a biochemical property, a reaction to therapy, or survival. Interestingly, clinicians tend to emphasize the clinical outcome as the best variable (*21*). From the proposed gold standards, strictly diagnostic systems such as PNEUMON-IA can only employ the expert's opinion.

To compare expert system opinions with clinical experts judgements a "grad-

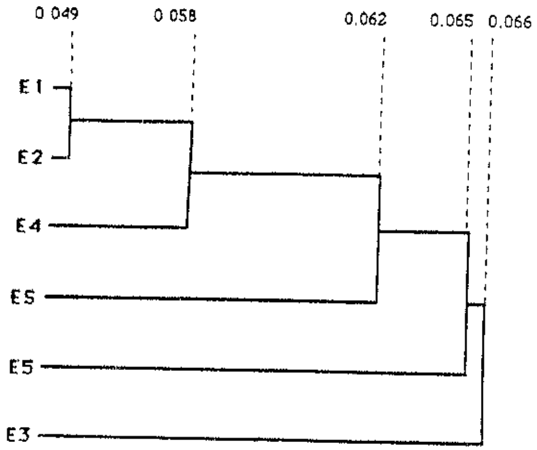

FIG. 7. Dendrogram of cluster analyses from city-block distances.

ing approach'' can be used. Briefly, it offers medical records and the diagnosis given by the expert system to a panel of expert clinicians which "grades" the expert system conclusions. Typical appraisal reports of accuracy include a list of degrees and the proportion of expert system diagnosis which merit each one (i.e., unacceptable: 5%, acceptable: 70%, ideal: 25%). Limitations inherent to this approach rely on the subjectivity of the grading process and the lack of a real

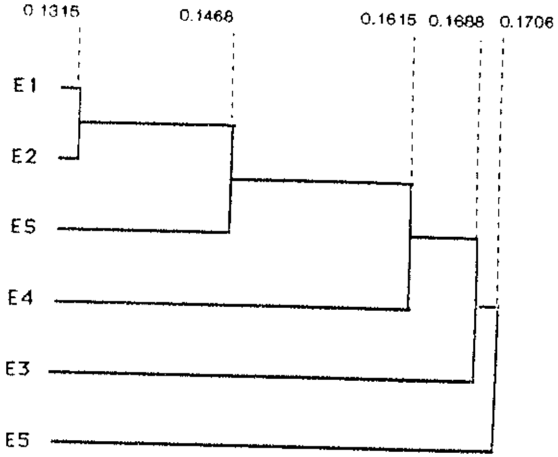

FIG. 8. Dendrogram of cluster analyses from euclidean distances.

"human vs machine" comparison. To partially overcome the latter, diagnosis emitted by clinicians on the same cases can be concurrently graded (*10*).

Under a methodological point of view, the system used to validate PNEUMON-IA is a new approach as it evaluates the expert system from different perspectives, using two groups of techniques: frequency analysis and distance analysis. Frequency analysis includes the contrast between the number of diagnosis given by clinicians and PNEUMON-IA and the number of omitted and singular etiologies. All these techniques are interdependent and give, among other data, information about diagnostic exhaustivity.

Distance analysis supplies also some degree of "grading" of the expert system among clinicians (as well as of clinicians among themselves). The approach used to test PNEUMON-IA combines simplicity and comprehensiveness, and can be helpful to validate any other system with similar characteristics.

An alternative and widely used tool for assessing interrater (interexpert) reliability is the Cohen kappa coefficient. Several formulations of this coefficient have been proposed: from the simplest that assesses the agreement between two raters (*22*), to the case where the reliability among multiple raters using nominal variables is considered (*23*), to the more complex case where each rater of a group chooses a subset of items from a list (i.e., a subset of possible diagnostics for a patient) (*24*). Our case is still more complex: several raters using an ordinal variable to qualify the possibility of each diagnostic of a set. Although it could be theoretically possible to generalize the kappa coefficient to our problem and compare the results obtained with each strategy, the exercise is beyond the scope of this paper.

The results obtained applying our approach to PNEUMON-IA validation show that differences between etiological diagnoses made by the expert system and those made by some specialists were smaller than differences between some specialists themselves. Specialists with highest proficiency scores gave the closest diagnoses. The best specialist was, moreover, the one with the least omitted etiologies and least singular etiologies. The expert system was closer to the "best" than to the "worst" specialist. PNEUMON-IA supplied the highest number of singular etiologies, largely because it takes viral etiologies into consideration more often. The success of the expert system in cases of confirmed etiologies was similar to that of clinical specialists. Etiological diagnoses emitted by the expert system PNEUMON-IA agreed with the best known specialists in our area.

In summary, in a knowledge area where a gold standard is not readily available, this method of validation gives a broad overview of the system expertise.

APPENDIX

EXPERT SYSTEM	PNEUMON-IA Knowledge Base	Date: 03-01-89
IDENTIFICATION	Clinical Record : Joan H	Hospital: La Alianca

FACTS OBSERVED

FACT	Internal_Name	Value
1 Community Acquired Pneumonia	NE	Yes
2 Severity	ESTAT_MALALT	Light
3 Coma	COMA	No
4 Obnubilation	DESORIENTAT	No
5 Days from onset of symptoms	DIES_INICI	10
6 Days from Diagnostic	DIES	1
7 Day	DIA	1
8 Month	MES	10
9 Year	ANY	88
10 Age	EDAT	19
11 Sex	SEXE	Man
12 Past history significant	AP_INTEREST	No
13 Previous antibiotics	ABS	No
14 Onset form	INICI	Subacute
15 Cough	TOS	Yes
16 Expectoration	EXPECTORACIO	Yes
17 Sputum	ESPUT	Mucous
18 Pleuritic pain	DOLOR_PLEURITIC	Yes
19 Fever	TEMPERATURA	37.5
20 Heart rate	FRE_CAR	60
21 Consolidation signs	SIGNES_CON	UK/UC
22 Condensation signs	SEMIOL_CON	No
23 Associated clinical data	TE_ASSOC	Yes
24 Kind of Associated Clin Data	ASSOCIA_GRA	Rheumatological Gastrointest Respiratory
25 G.I. Data	ASSOCIA_GI	Diarrhoea Abd. Distension
26 Rheumatology Data	ASSOCIA_REU	Join pain Myalgia
27 Respiratory Data	ASSOCIA_RES	Cold
28 RX exists	RX_FET	Yes
29 RX Data	DADES_RX	Interstitial
30 Certainty Interstitial Pat.	RX_INTER_S	Quite Possible
31 Distribution of Infiltrate	RX_DISTR	Segmentary
32 Certainty Segmentary	SEG_S	Quite Possible
33 Multisegmentary	RX_SEG_VAA	No
34 Localization of Infiltrate	LOCALITZACIO_S	Mid_lobe
35 Laboratory	ANALAB	Yes
36 Number of Leucocytes	LEUCOCITS	9000
37 Number of Polinucleated	POLIS	6300
38 Percentage of Polinucleated	PP	70
39 Number of bands	BANDA	0
40 Evolution	EVOLUCIO	Acute
41 Headache	MAL_CAP	Yes
42 Contact with pneumonia	ALTREPN	No
43 Coombs done	FET_COOMBS	No
44 Crioaglutinines done	FET_CRIO	No
45 Birds in house	OCELL	No
46 Perihiliar	PERIH	Very Possible

FACTS INFERRED (No Diagnostical)

1 Neutrofilia	NEUTROFILIA	Almost Impossible
2 Left Deviation	DESV_ESQ	Almost Impossible
3 Bacterial infection	BACT	Moderately Possible
4 Atipic Pneumonia	ATIP	Very Possible

SYSTEM DIAGNOSIS

1	Viral	Very Possible
2	Mycoplasma	Almost Sure
3	Chlamydia	Slightly Possible

RULES USED

R20001	R20002	R2005	R20018	R20020	R18012	R17009
R17010	R17011	R17013	R17016	R03009	R03014	R03024
R03026	R03028	R01001	R01007	R01008	R01009	
R01009A	R01009B	R01009C	R01028	R01029	R01030	R01057
R01071	R01078	R01094	R01101			

METARULES USED

R25892A R25893 R25898 R25896 R25602

Acknowledgment

We thank Marta Pulido, M.D., for editorial assistance.

References

1. López de Mántaras, R., Cortés, U., Sierra, C., Villar, A., and Plaza, E. An expert system building tool with approximate reasoning. *In* "Fuzzy Logic in Knowledge Engineering" (C. V. Negoitia and H. Prade, Eds.), Rheinhold Interdisciplinary System Research Serie No. 86, pp. 182–193. Verlag, Tür, 1988.
2. Miller, P. L. Evaluation of artificial intelligence systems in medicine. *In* "Selected Topics in Medical Artificial Intelligence" (P. L. Miller, Ed.). Springer-Verlag, New York, 1988.
3. Verdaguer, A. PNEUMON-IA: desenvolupament d'un sistema expert d'ajuda al diagnòstic mèdic. Doctoral thesis. Universitat Autònoma de Barcelona, Barcelona, 1989.
4. Godó, Ll., López Mántaras, R., Sierra, C., and Verdaguer, A. Managing linguistically expressed uncertainty in MILORD application to medical diagnosis. *Artif. Intel. Commun.* **1**, 14 (1988).
5. Godó, Ll., López Mántaras, R., Sierra, C., and Verdaguer, A. MILORD: The architecture and management of linguistically expressed uncertainty. *Int. J. Intel. Sys.* **4**, 471 (1989).
6. Yu, V. L., Fagan, L., Bennet, S. W., Clancey, W. J., Scott, A. C., Hannigan, J. F., Buchanan, B. G., and Cohen, S. N. An evaluation of MYCIN's advice. *In* "Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project" (B. G. Buchanan and S. N. Cohen, Eds.) pp. 589–596. Addison–Wesley, Reading, MA, 1984.
7. Kingsland, L. C. III. Evaluation of expert systems. Experience with the AI/Rheum knowledge-based consultant system in rheumatology. *In* "Selected Topics in Medical Artificial Intelligence" (P. L. Miller, Ed.). Springer-Verlag, New York, 1988.
8. Reggia, J. A., Tabb, D. R., Price, T. R., Banko, M., and Hebel, R. Computed-aided assessment of transient ischemic attacks: A clinical evaluation. *Arch. Neurol.* **41**, 1248 (1984).
9. Soula, G., Vialettes, B., and San Marco, J. Une validation multicentrique du système expert PROTIS. *In* "AAVV: Proceedings, 7th International Symposium on Expert Systems and Applications," pp. 239–255. Avignon, 1987.
10. Quaglini, S., Stefanelli, M., Barosi, G., and Berzuini, A. A performance evaluation of the expert system ANEMIA. *Comput. Biomed. Res.* **21**, 307–323 (1988).
11. Wite, R. J., Blainey, A. D., Harrison, K.J., and Clarke, S. K. R. Causes of pneumonia presenting to a district general hospital. *Thorax* **36**, 566 (1981).
12. MacFarlane, J. T., Finch, R. G., Ward, M. J., and Macrae, A. D. Hospital study of adult community-acquired pneumonia. *Lancet* **2**, 255 (1982).
13. Almirante, B., Pahissa, A., Planes, A. M., and Martínez-Vázquez, J. M. Etiología de las neumonías extrahospitalarias. *Med. Clin. (Barcelona)* **85**, 175 (1985).
14. Berntsson, E., Blomberg, J., Lagergard, T., and Trollfors, B. Etiology of community-acquired pneumonia in patients requiring hospitalization. *Eur. J. Clin. Microbiol.* **4**, 268–272 (1985).
15. Marrie, T. J., Garyston, J. T., Wang, S., and Kuo, C. Pneumonia associated with TWAR stain of Chlamydia. *Ann Intern. Med.* **106**, 507 (1987).
16. Ausina, V., Coll, P., Sambeat, M., Puig, I., Condom, M. J., Luquin, M., Ballester, F., and Prats, G. Prospective study on the etiology of community-acquired pneumonia in children and adults in Spain. *Eur. J. Clin. Microbiol. Infect. Dis.***14**, 343–347 (1988).
17. Norusis, M. J. "SPSSx Advanced Statistics Guide." SPSS, Chicago, 1985.
18. Sutton, G. C. How accurate is computer-aided diagnosis? *Lancet* **2**, 905–908 (1989).
19. Chard, T. Human versus machine: A comparison of a computer 'Expert System' with human experts in the diagnosis of vaginal discharge. *Int. J. BioMed. Comput.* **20**, 71–78 (1987).
20. Sutton, G. C. Computer aided diagnosis: A review. *Br. J. Surg.* **76**, 82–85 (1989).

21. BAAK, J. P.A. Artificial intelligence systems (expert systems) as diagnostic consultants for the cytologic and histologic diagnosis of cancer. *J. Cancer Res. Clin. Oncol.* **114,** 325–334 (1988).
22. KRAEMER, H. C. Extension of the Kappa coefficient. *Biometrics* **36,** 207–216 (1980).
23. POSNER, K. L., SAMPSON, P. D., CAPLAN, R. A., WARD, R. J., AND CHENEY F. W. Measuring interrater reliability among multiple raters: An example of methods for nominal data. *Stat. Med.* 9, 1103–1115 (1990).
24. AHN, C. H., AND MEZZICH, J. E. PROPOV-K: A FORTRAN program for computing a Kappa coefficient using a proportional overlap procedure. *Comput. Biomed. Res.* **22,** 415–423 (1989).

COMPUTERS AND BIOMEDICAL RESEARCH **25,** 527–537 (1992)

Plots for Examination of Univariate Twin Data

CHRISTOPHER J. WILLIAMS*,†,[1] AND JOE C. CHRISTIAN*

**Department of Medical and Molecular Genetics and †Division of Biostatistics, Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana*

Received February 25, 1992

We discuss informative plots for univariate twin data that can be used in conjunction with twin data analyses. The plots are useful for spotting outliers, spotting possible single gene effects, and displaying the contribution of individual twin pairs to the fit of genetic models of the data. We illustrate the use of the plots on bone mineral data, and present programs for generating the plots in SAS.

I. INTRODUCTION

The ability of researchers to conduct sophisticated statistical analyses of twin data is rapidly increasing. This is due to the introduction of flexible models for twin data analysis (*1, 2*), to increased availability of computer programs to perform these analyses (*3, 4*), and to tremendous recent advances in computer hardware. The increased emphasis of fitting more interesting models to data, however, at times appears to overshadow the role of simple exploratory data methods such as plots. The role of visual displays of data was perhaps best described by Sir. R. A. Fisher:

> The preliminary examination of most data is facilitated by the use of diagrams. Diagrams prove nothing, but bring outstanding features readily to the eye; they are therefore no substitute for such critical tests as may be applied to the data, but are valuable in suggesting such tests, and in explaining the conclusions founded upon them. (*5*)

In this spirit we present two simple plots for examination of univariate twin data. We suggest that plots such as these be routinely examined, and further that presentation of these plots along with twin analyses greatly enhances understanding of results. We illustrate the use of the plots on bone study data, and we present a macro for producing these plots in the Statistical Analysis System computer package (SAS Institute, Cary, NC 27511).

[1] To whom reprint requests should be addressed at: Department of Mathematics and Statistics, University of Idaho, Moscow, ID 83843.

0010-4809/92 $5.00