The New Navigators: from Professionals to Patients
R. Baud et al. (Eds.)
IOS Press, 2003

Watching symptoms and illnesses through the eyes of Multiple Correspondence Analysis: a case study in Toxicology

Vincent Rialle[a], Vincent Danel[b]

[a]*Laboratoire TIMC-IMAG, Grenoble, France*
[b]*Service de Toxicologie Clinique, Hôpital Michallon, Grenoble, France*

Abstract

The goal of the reported research was the design of a computerized tool aimed at aiding an emergency specialist, in a toxicological emergency unit, to quickly identify, at the admission time, the various poisons ingested by a comatose patient. This medical decision making problem has been proved to be computationally intractable since a lot of patient cases sharing the same clinical table have different diagnoses in terms of psychotropes combinations. The paper explores the idea that the outcome of Multiple Correspondence Analysis (MCA), a mathematical data analysis method, can be thought of as visual and analytical aids for the physician facing this decision problem. We argue that the expert's clinical reasoning can be enhanced by the factorial maps and some computerized results provided by MCA. Using a learning database of 505 diagnosed cases, we realized a whole decision aiding system called TOXSYMEDIA. A test-base of 97 patients was used to partially assess the system. The method and resulting tool revealed to be appreciable to early inform the physician about the possible combination of ingested psychotropes.

Keywords:
Medical Decision Making; Clinical Toxicology; Psychotrope Induced coma; Multiple Correspondence Analysis

1. Introduction

The goal of the reported research was the design of a computerized tool aimed at aiding an emergency specialist, in a toxicological emergency unit, to quickly identify, at the admission time, the various poisons ingested by a comatose patient. Acute poisoning is one of the most frequent emergencies in developed countries, and suicide attempts with pharmaceutical drugs are by far the main circumstance. Rigorous and exhaustive identification of the whole set of toxic causes of the coma is possible only by means of biological and toxicological analyses. Unfortunately, these analyses require an important delay ranging from half an hour to several days, depending on the type of analyze, whereas the outcome of the treatment essentially depends on the primary emergency care. On the other hand, no explicit knowledge relating clinical findings to poison combinations is available so far. This lack of knowledge is mainly due to the increasing number of psychotropes and their occurring ingested combinations, along with the weakness of clinical findings to discriminate them. In the point of view of medical decision making, a first difficulty lies in the lack of explicit knowledge concerning multipoisoning. The only source of reliable knowledge is the database of patient records gathering rigorously examined and diagnosed cases. A second difficulty is that a good number of cases share the same clinical findings while having different diagnoses in terms of psychotropes combinations. This feature renders any classical machine learning solution unable to provide any acceptable classifier. The decisional problem of psychotrope induced coma

(PIC) cannot be solved by sheer logical reasoning because of local contradictions: same clinical findings, different Toxicological Diagnosis (TD). On the other hand, the use of uncertain reasoning such as probabilistic or fuzzy logic, decision trees, neural networks or genetic algorithms, results in non-profitable responses due to a very high level of uncertainty. This is why we explored the idea that graphical and numerical outcomes of Multiple Correspondence Analysis (MCA) could provide some valuable and be though of as visual and analytical cognitive extensions of the physician when no other computerized decision aiding method can be properly used.

2. Materials and Methods

2.1. Medical decision making issue and experimental input data

PIC is characterised by a lack of neurological focal signs. Pupils are symmetric, equal in size and reactive to light, even in very deep coma. Neurological symptoms such as muscular tonus, tendon reflexes, and pupil size, may vary according to the ingested substance. Some psychotropes can lead to cardiac complications. Dysrhythmias are mainly due to a membrane stabilising effect characterised by a wide QRS complex and a prolonged QT interval on the ECG. Urine retention may be observed in tricyclic antidepressants and phenothiazines ingestion. No single symptom is specific of only one substance, only a group of symptoms (syndrome) can lead to a diagnosis.

A database of 505 patient cases concerned exclusively with the psychotropes ingestion with or without alcohol was used as 'learning base' to compute de factorial maps of the (MCA) (cf. 2.3). The database featured 13 input variables (mainly neurological and cardio-vascular) (Table 1), and 28 psychotropes combinations (Table 2). Input data was characterised by their facility to be observed in emergency, their capability of describing a PIC, and their statistical relevance for identifying the psychotropes combinations [1]. None of the selected had a significant contribution to the discrimination of TDs.

Table 1 – Clinical input data

Nb	*Input data*	*Set of values*
1	Patient sex	F / M
2	Temperature	low / normal / high
3	Light reflexes	quiet / restless
4	Photomotors	present / absent
5	Eyes position	normal / abnormal
6	Pupils	myosis/interm. / mydriasis
7	Tonus	hypertonia / hypotonia
8	Tendon reflexes	brisk / diminished
9	Systol. blood pressure	low / normal / high
10	Pulse rate	low / normal / high
11	QRS complex	normal / prolonged
12	Qt interval	normal / prolonged
13	Urine retention	yes / no

Table 2 – Names and codes of individual psychotropes

Alcohol	E
ADT : Tricyclic antidepressants (TCA)	a
Benzodiazepines	b
Barbiturates	B
Carbamates	c
Morphine	m
Phenothiazines	p

2.2. Role and Features of Multiple Correspondence Analysis

MCA is a mathematical data analysis method based on the calculus of best projection plans in the space of input data (or *variables*) describing the problem [2]. It uses only categorical variables, and its main difference with Principal Component Analysis is that input data and output decisions can be represented in the same visual bi-dimensional maps. Furthermore, MCA is peculiarly interesting for its ability to model non-linear relationships [2, 3]. It consists of a series of transformations applied to a so called 'disjunctive matrix' (or D-matrix) that represents the patient coordinates in the vector space of Boolean variables given by variable values (Table 3).

Table 3 – Fragment of the D-matrix

case	Sex= F	Sex= M	Temp= Low	Temp= normal	Temp= high	State= quiet	State= agitated	etc.	TD
1	1	0	0	0	0	1	0	...	EBC.
2	0	1	1	0	0	1	0	...	CP
3	1	0	1	0	0	1	0	...	bBC
etc.									

The successive transformations of the D-matrix lead to a 'matrix of profiles' in which each element represents a relative frequency. Afterwards, a complex algorithm that could not be easily explain in a few words (see [3] for a detailed presentation), is triggered to compute the factorial axes and maps, along with the TD and patient dot coordinates. Distance between dots, on factorial maps, can be computed and displayed as additional information. However, this distance cannot be interpreted as a straightforward Euclidean distance since its interpretation depends on the relative position of the dots according to some interpretation rules (see next section for examples or rules).

The constraint of categorical variables imposed by MCA methodology has not been a limitation since the only 4 quantitative input data (Temperature, Systolic blood pressure, Pulse rate, QRS complex) is systematically interpreted by the physician as categorical variables by means of clinically usual segmentation rules.

3. Results

3.1. Decision aiding system

We realized a whole decision aiding system called TOXSYMEDIA (for TOXicology application of the concept of SYnergy between Medicine and Artificial Intelligence) using Matlab™ as programming environment, and the learning base described above. The system has been designed to provide the clinician with a broad range of facilities - by means of a user friendly interface (Fig. 1) -, regarding the access to databases, maps, interpretations of dots, and evaluation. The following are the main features: 1) displaying patient dots for a chosen TD; 2) displaying quality of the projections (cosines of the angle between the dots and the plan); 3) merging of a new set of pre-classified cases with the current learning case-base; 4) computing eigen values and eigen vectors for factorial maps; 5) giving various on-line helps and information about the user interface, the way of interpreting plans, etc.; 6) testing the performance of the classification by means of the Euclidean distance.

Here is an example to illustrate how the system works in relation with the physician: the system is given the 13 findings shown on table 4, by means of the graphical user interface

(Fig. 1), and computes and displays the first three MCA factorial maps (Fig. 2, 3, 4).

Table 4 - Example of patient case

1.	Sex	:	M
2.	Temperat.	:	37° C
3.	Light refl.	:	restless
4.	Photom.	:	present
5.	Eyes pos.	:	normal
6.	Pupils	:	mydriasis
7.	Tonus	:	hypertonia
8.	Tendon r.	:	brisk
9.	S.B.P.	:	100mmHg (high)
10.	Pulse rate	:	124 c/min (high)
11.	QRS complex	:	0,08 s (prolong.)
12.	Qt interval	:	prolonged
13.	Urine ret.	:	No

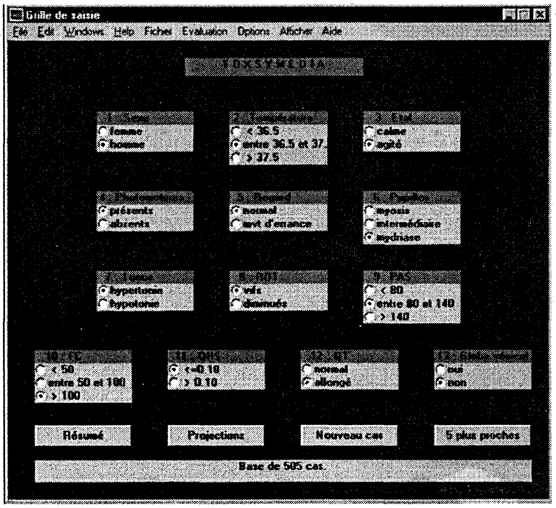

Figure 1 – Graphical user interface for clinical data input (in french)

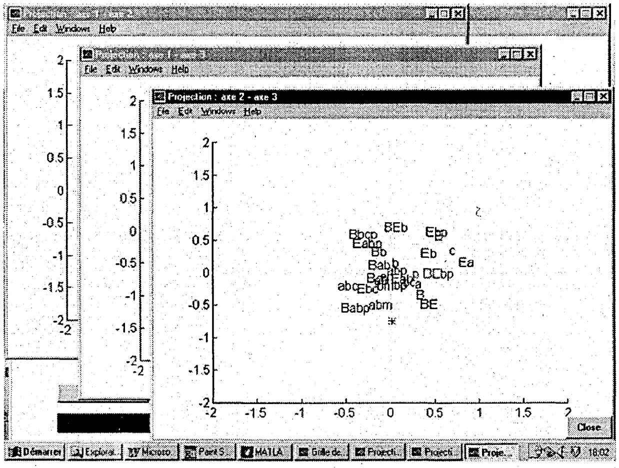

Figure 2 - MCA factorial plans

Each label displayed on the maps (*e.g.* Babp, abm, BE...) is a TD. Its location represents the coordinates of the barycentre of set of patient cases having this TD. For instance, the label *ab* on the maps (Fig. 2, 3, 4) stands for the combination of 'antidepressants' and 'barbiturate' (Table 2), and represents the mean point, or barycentre, of the 104 cases, among the 505 cases of the learning base, that have 'antidepressants and barbiturate' as TD. In the same way, the 'carbamates' TD is represented by the dot labelled *c* on the maps, which corresponds to the barycentre of the cases featuring 'carbamates' as TD. The star on the factorial maps represents the patient case under study. A Euclidean distance between its coordinates and the five closest diagnoses occurring in the learning database is also displayed (Fig. 4). One can observe that the visually closest TD of the exemplar patient are *a*, *ap*, and *Bcp*. The interpretation of this graphical information is: - there is a strong indication in favour of tricyclic antidepressants and phenothiazines; a weaker indication for tricyclic antidepressants alone; an even weaker indication for the combination of benzodiazepines, carbamates, and phenothiazines. Afterwards, the physician can compare these "visual closeness" with the list of five closest TD dots according to the Euclidean distance (Fig. 4).

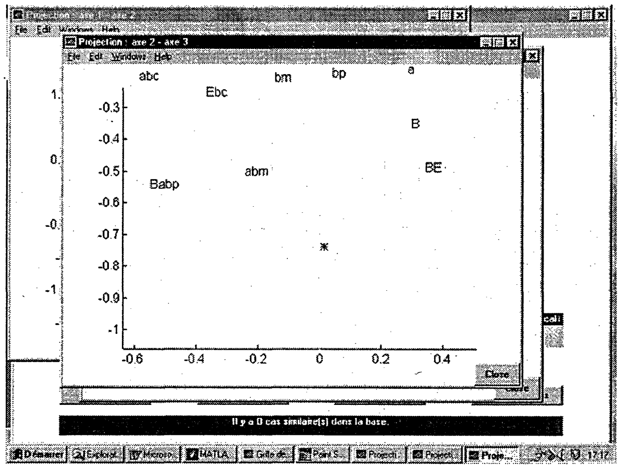

Figure 3 - Zooming/re-sizing facilities

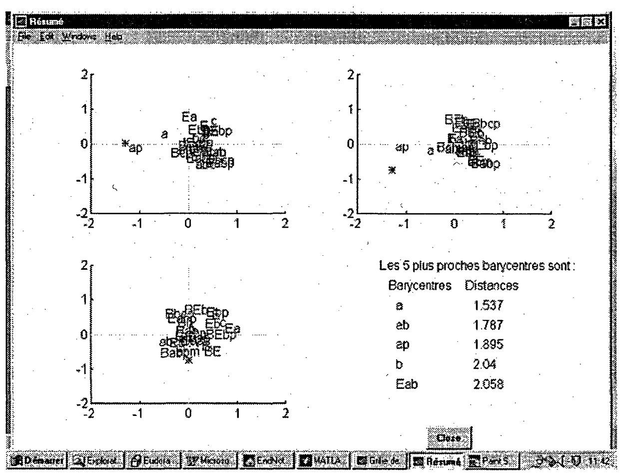

Figure 4 - Summarizing display

The factorial maps can be used as a starting point for exploring further relationships. However, interpreting them requires a few guideline rules such as: - keep in mind that the interpretation of the location of displayed TD must be pondered with the percentage of variance captured by the map; - (as a consequence) need for several maps, since each map features only a certain percentage of the whole input space information (represented by the variance), and is orthogonal to all other ones. In the coma application, the three first maps feature about 80 % of the global variance, which is quite commonplace; - keep in mind the 'effect of perspective' of the projections (*e.g.* two dots which appear to be close on a map can be much more distant on another one); - two close dots reveal a strong relation between them, and the farthest two close dots are from the axes origin, the more they are correlated.

So, in the course of experiments, the factorial maps have revealed to be a valuable means to graphically summarise the relations between toxicological diagnoses and a peculiar patient's description. The maps have been useful to evaluate the most likely diagnostic possibilities.

3.2. Evaluation

A kind of subjective help is provided to the physician by the labelled dots on the factorial maps. Though this help is very difficult to assess, it but must be taken into account for the appreciation of the system. However, we tried to quantitatively evaluate the relevance of the system in terms of diagnosis accuracy. To that end, a test-base of 97 patients with TD given by lab results was realized to assess the system.

The following test was designed to compare the actual diagnosis given by the test-base with the one proposed by the system: diagnosis accuracy was considered to be good when the actual diagnosis was one of the three first TD proposed by the system (as shown on Fig. 4). This test was positive for 40 % of the 97 cases of the test-base. This weak percentage of good responses was expected because of the intractability of the problem.

Due to the lack of prior knowledge about symptoms of multiple psychotropes-induced comas, we could not compare this result with any previously existing knowledge.

An increase of the number of cases in the database should not necessarily improve the discrimination between the different diagnoses; however, we do think that it should improve the performance of MCA used as a classifier by means of the above described test.

4. Conclusion

This study had no other ambition than exploring a way to enhance the emergency care of comatose patients by ingestion of several different psychotropes. The goal was to find a way for anticipating the lab-results whose delay can be disastrous. We showed that the clinical response to the search for the combination of ingested psychotropes is not supported by clinical rules, and that the attempt to find a solution with machine learning algorithms also failed. So, the explored solution has been a visual representation of the clinical situation supported by a rigorous mathematical data analysis approach: the MCA. This method revealed to be an appreciable tool to early inform the physician about the possible combination of psychotropes.

Acknowledgments

This research was supported by the University Hospital Centre of Grenoble and by the French Ministry of Research ("GIS Sciences de la Cognition").

References

[1] Danel V, Cony P, Rialle V. Coma due to psychotropes. How can multifactorial analysis help make the diagnosis. Proc. "Computer as an aid in Poison Centres", 2nd meeting of the European Association of Poison Centres and Clinical Toxicologists, 1995; Lille; pp. 53.

[2] Benzécri J-P, Benzécri F. *Analyse des Correspondances : exposé élémentaire.* Paris: Dunod, 1980.

[3] Anderson TW, Finn JD. *The new statistical analysis of data.* New York: Springer, 1996.

Address for correspondence

V. Rialle, Laboratoire TIMC-IMAG, Faculté de Médecine de Grenoble, 38700 La Tronche, France.
Mail :Vincent.Rialle@imag.fr