Non-linear associations between laryngo-pharyngeal symptoms of gastro-oesophageal reflux disease: clues from artificial intelligence analysis

Associazioni non lineari tra sintomi laringo-faringei della MRGE: indizi dalle analisi basate sull'intelligenza artificiale

E. GROSSI
Medical Department, Bracco SpA, Milano, Italy

Key words

GERD • Laryngo-pharyngeal symptoms • Non-linear association • Artificial intelligence

Parole chiave

GERD • Sintomi laringofaringei • Associazione non lineare • Intelligenza artificiale

Summary

The relationship between the different symptoms of gastro-oesophageal reflux disease remain markedly obscure due to the high underlying non-linearity and the lack of studies focusing on the problem. Aim of this study was to evaluate the hidden relationships between the triad of symptoms related to gastro-oesophageal reflux disease using advanced mathematical techniques, borrowed from the artificial intelligence field, in a cohort of patients with oesophagitis. A total of 388 patients (from 60 centres) with endoscopic evidence of oesophagitis were recruited. The severity of oesophagitis was scored by means of the Savary-Miller classification. PST algorithm was employed. This study shows that laryngo-pharyngeal symptoms related to gastro-oesophageal reflux disease are correlated even if in a non-linear way.

Riassunto

Le relazioni tra i differenti sintomi del reflusso gastroesofageo (GERD) rimangono ancora da chiarire poiché non esistono studi confrontabili. Scopo del nostro studio è stato quello di studiare i vari sintomi del GERD e quelli faringolaringei in particolare utilizzando un modello matematico di analisi già impiegato nelle tecniche di intelligenza artificiale (algoritmo PST). Sono stati studiati 388 pazienti provenienti da 60 Centri Clinici di riferimento affetti da esofagite e stadiati con la classificazione di Savary-Miller. Questo studio dimostra che i sintomi laringofaringei se associati si correlano bene nel GERD anche se in misura non lineare.

Introduction

Gastro-oesophageal reflux disease (GERD) has recently been defined as the presence of oesophageal mucosal interruptions or the occurrence of reflux-induced symptoms severe enough to significantly impair quality of life. The importance of symptom evaluation and assessment is already evident from the definition in the management of the disease. Despite the fact that symptom analysis is considered to be important to determine whether reflux-induced symptoms are sufficiently severe to justify the diagnosis of reflux disease, there is no reliable method to use this kind of information as a diagnostic tool, as a result of the non-linearity of the relationship between the quality and intensity of symptoms and the diagnostic target.

The relationship between the different symptoms of GERD remain markedly obscure due to the high underlying non-linearity and the lack of suitable studies focusing on the problem. In particular, as far as concerns atypical symptoms such as those affecting the respiratory tract or the larynx, their deep structure and clustering remain to be fully elucidated.

Recently, it has been emphasized that a specific triad of laryngo-pharyngeal symptoms (hoarseness, lump in the throat and cough) tends to appear in a subset of patients affected by GERD, even if the consistency of this association is rather weak from a statistical point of view [1].

The aim of this study was to evaluate the hidden relationships between this triad of symptoms related to GERD using advanced mathematical techniques, borrowed from the artificial intelligence field, in a cohort of patients with oesophagitis.

Material and methods

Data base

The data for this analysis come from a large observational study on GERD patients (with and without erosive oesophagitis at endoscopy) diagnosed by means

Table I. Variables recorded in study group.

Helicobacter positivity	Food ascent
Helicobacter negativity	Lump in throat
Heartburn occurrence	Belching
Heartburn frequency	Vomiting
Heartburn severity	Vomiting frequency
Regurgitation occurrence	Hematemesis
Regurgitation frequency	Hiccup
Regurgitation severity	Anorexia
Night regurgitation	Cough
Chest pain frequency	Wheezing
Chest pain severity	Shortness of breath
Swallowing defect	Dyspnea
Swallowing defect severity	Hoarseness
Swallowing defect frequency	Head up
Swallowing defect type	Smoking
Epigastric pain	Coffee
	Alcohol

of endoscopy and a positive proton pump inhibitor (PPI) test [2].

A total of 634 patients (311 females, mean age 43.5 ± 12.4 (SD) years) with typical GERD symptoms were observed by 60 Italian Gastroenterology Units during 2003-2005; patients were given a previously validated structured questionnaire QuID systematically assessing all typical and atypical symptoms related to GERD according to occurrence, frequency and severity.

For the present study, 388 patients with endoscopic evidence of oesophagitis were recruited. The severity of oesophagitis, scored by means of Savary-Miller classification, was as follows: grade 1 = 75.4%, grade 2 = 18.2%, grade 3 = 3.5% and grade 4 = 2.9%. Overall, 33 variables were recorded, most

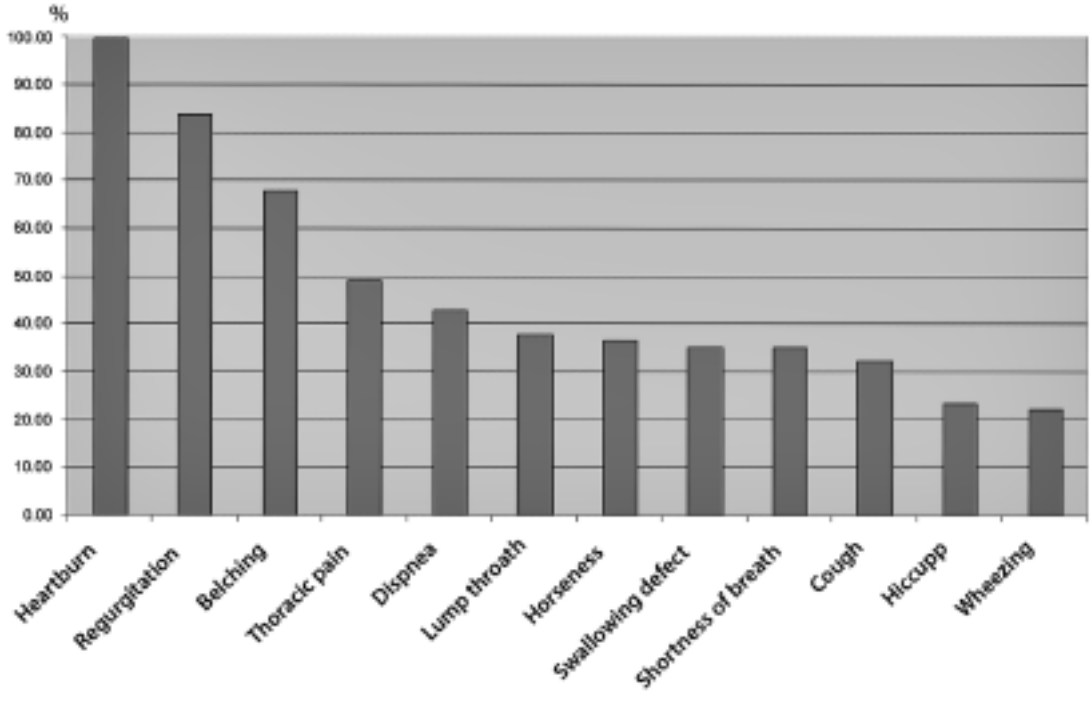


Fig. 1. Prevalence of typical and atypical symptoms associated with GERD

Table II. Linear correlation between ENT triad symptoms.

	R	r^2
Lump cough	0.13	0.0169
Lump hoarseness	0.22	0.0484
Cough hoarseness	0.2	0.04

of which related to typical and atypical symptoms (Table I).

Descriptive analysis

The prevalence of typical and atypical symptoms associated with GERD in the 388 patients shown in Figure 1 reveals that approximately one third of the patients reported at least one respiratory or laryngo-pharyngeal symptom.
The simultaneous occurrence of atypical symptoms was infrequent: 13 patients reported the simultaneous presence of respiratory and laryngo-pharyngeal symptoms (lump in throat, cough, wheezing, shortness of breath, dyspnea, and hoarseness), and 37 patients reported the simultaneous presence of hoarseness, lump in the throat and cough.
Statistical analysis did not reveal any associations between symptom clusters due to the poor linear correlation.
In particular, the matrix of correlation between lump in the throat, cough and hoarseness showed modest values of r2 (Table II).

Buscema at the Semeion Research Center, in 1999, which aims to underline factors that constitute the basic structure of observed data placing all points of a dataset onto a 2D space using an evolutionary algorithm [3] and minimizing the distortion of original distances between points.
PST approximates the solution, without knowing if it exists and without *a priori* knowledge of research space structure.
Every matrix of vectorial distances can be projected onto a D-dimensional map.
The problem: defining a Map Distance:

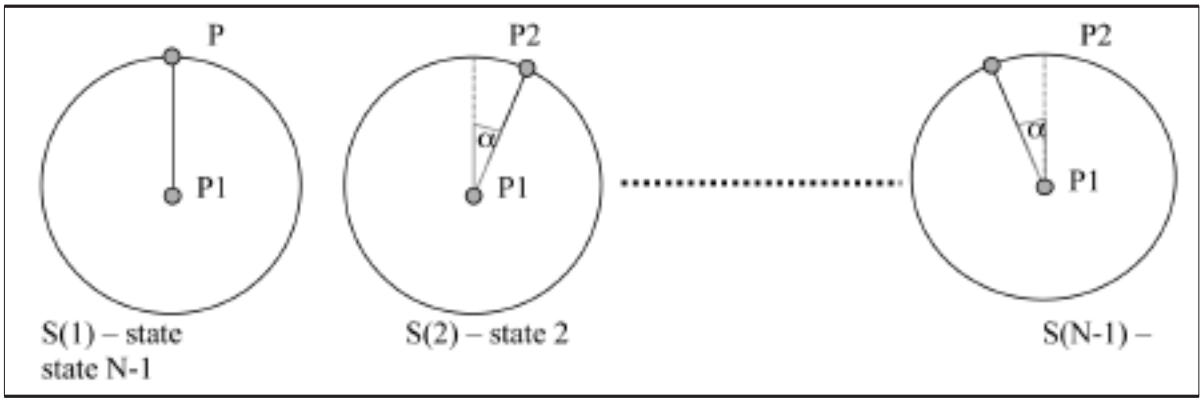
$$Md_{ij} = \sqrt[2]{(Px_i - Px_j)^2 + (Py_i - Py_j)^2}$$

and a Vector Distance $Vd_{ij} = \sum_{k=1}^{L} \left| Pv_{ik} - Pv_{jk} \right|$

we can set up the optimization problem:

$$\text{min E;} \quad E = \frac{1}{C} \cdot \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left| Md_{ij} - Vd_{ij} \right|; \quad C = \frac{N \cdot (N-1)}{2}$$

We define: State, S, a configuration of points on a plane with known distances between them, despite the rotation of the configuration.
Angle of tolerance, α, the angle defining an arc on the circumference where two points are not distinct.



Since one of the limitations of statistical procedures aiming to establish the association between variables is related to the linearity of relationship between variables, we approached the problem through a non-linear analysis tool: the PST algorithm.

PST algorithm

PST is an evolutionary architecture, conceived by M.

The number of possible states starting from the distances between N points is:

$$S = \alpha^{(N-2)}$$

Define P as the number of tests necessary to verify the distances between N points in one state:

Table III.

	LA	NY	Boston	Detroit	Buffalo	Pittsburg	Chicago	Saint Louis	Cincinnati	Dallas	Atlanta	Memphis
LA	0											
NY	5600	0										
Boston	6109	509	0									
Detroit	4582	1145	1527	0								
Buffalo	5091	764	1018	509	0							
Pittsburgh	4836	764	1145	509	382	0						
Chicago	4073	1655	2036	509	1018	891	0					
Saint Louis	3564	2063	2418	1018	1527	1273	636	0				
Cincinnati	4327	1273	1655	382	764	509	509	764	0			
Dallas	2800	2927	3436	2036	2545	2291	1655	1018	1782	0		
Atlanta	4327	1527	2036	1145	1400	1018	1145	1018	164	1527	0	
Memphis	3564	2164	2545	1273	1782	1400	1018	382	891	891	764	0

Flight Distances in a geographic space between 12 USA Cities (in miles)

$$P = \frac{M \cdot (M-1)}{2}; \qquad M = N - 2.$$

Total number of tests, in all possible states: $Q = S \cdot P$

The mapping problem we have presented is a NP problem, highly complex when the angle of tolerance, expressed in radians, increases (Table III).

Every air route has three types of alteration in a 2D Euclidean space:
1) a longitudinal alteration;
2) an altitude alteration;
3) a structural alteration.

Two dimensional map created by PST: it places the 12 US cities in a very similar way to how they actually appear on physical maps.

Results

The bidimensional map of the variables under study is shown in Figure 2.
In this plot, the degree of proximity between the variables is proportional to the degree of their non-linear association; the variables closer to each other are more associated. The different colours refer to the clusters membership. There are 10 variable clusters. It can be seen that the triad of interest has been attributed to the same cluster. This membership would have been missed with traditional statistical approaches such as principal component analysis.

Discussion

The most common algorithms of linear projections are the PCA (Principal Component Analysis) [4] and the ICA (Independent Component Analysis) [5]; the former requires a Gaussian distribution of data, while the latter does not require any specific distribution.
If the relationships between variables are non linear, the above-mentioned methods are not able to preserve, with sufficient accuracy, the geometrical structure of the original space. In the compression, in fact, much important information is lost. It is evident, then, the need both to eliminate the requirement of linearity between axes and to find a non-linear mapping between the original and the final space capable of preserving, as much as possible, the relationships between variables.
The approach shown to resolve the general problem of Topographic Mapping, presents two important elements: on the one hand, it emphasizes the advantage of tackling the problem from the optimization point of view, and on the other, it suggests the need to use a particularly efficient evolutionary algorithm to solve it.
The experimental results show, in an empirical way, that the space of solutions of the problem can be particularly complex and depends strongly on the data and on the function with which we measure the distortion of the original distances from those on the map. For this reason, the use of an algorithm, that has an excellent capacity to tackle different optimization problems and is not bound to specific properties of the problem itself, becomes fundamental.
GenD, the evolutionary algorithm used, shows the characteristic advantages of the algorithm of its class, elevated robustness on very different problems and, as pointed out, allows hybridization that increases, in a sensitive way, the velocity of convergence to the solution, traditional limit of the evolutionary algorithm.

Fig. 2. Non-linear mapping of the variables under study using the PST algorithm.

Conclusion

This study shows that laryngo-pharyngeal symptoms associated with GERD are, indeed, related.
Thanks to the PST algorithm it is, therefore, possible to conciliate experimental data with a common clinical experience according to which the triad described in this report represent a characteristic feature of a subset of GERD patients.

References

[1] Remacle M, Lawson G. *Diagnosis and management of laryngopharyngeal reflux disease*. Curr Opin Otolaryngol Head Neck Surg 2006;14:143-9.

[2] Caspani G, Intraligi M, Massini G, Capriotti M, Carlei V, Maurelli G, et al. *Progetto Esonet – Applicazione delle reti neurali artificiali nella diagnosi della MRGE*. Sistemi Artificiali Adattivi in Biomedicina 2005;2:118-57.

[3] Buscema M. *Genetic doping algorithm (GenD). Theory and applications*. Expert Systems 2004;21:63-79

[4] Hotelling H. *Analysis of a complex of statistical variables into principal components*. J Educat Psychol 1933;24:33-45

[5] Comon P. *Independent component analysis – a new concept?* Signal Processing 1994;36:287-314.

■ Address for correspondence: Dr. E. Grossi, Medical Department, Bracco SpA, via E. Folli 50, 20134 Milano, Italy. E-mail: enzo.grossi@bracco.com