Fuzzy clustering algorithm for multimorbidity patterns

Alba Ayala[1], *Maria João Forjaz*[2]

[1]aayala@isciii.es, National School of Public Health, Institute of Health Carlos III, Madrid, Spain

[2]jforjaz@isciii.es, National School of Public Health, Institute of Health Carlos III and REDISSEC, Madrid, Spain.

Abstract

This study aims to analyse the multimorbidity patterns in a sample of institutionalised older adults with dementia. Fuzzy clustering algorithm was applied to group a short list of diseases. Two clusters were identified. The fuzzy clustering method provides clusters that better reflect the complex structure of multimorbidity.

Keywords: fuzzy clustering, principal components, multimorbidity.

Introduction

Multimorbidity is understood as the simultaneous or sequential occurrence of two or more physical or mental disorders in the same patient. This concept has a different meanings and connotations that has been a source of disputes and conflicts [1].

Explanatory questions are relevant to the mechanisms underlying comorbidity and multimorbidity related to body and mind systems. For example, mental disorders are multifactorial, multidimensional and etiologically complex [1]. Therefore, explanatory models should refer mainly to pluralism instead of reductionism. The patterns of multimorbidity has been extensively researched in recent decades.

Multivariate methods may be used to find patterns of multimorbidity [2]. These methods study the interrelationships between diseases, facilitating the diagnosis and prevention and allowing obtaining care strategies for the most common combinations.

Fuzzy clustering algorithm is a multivariable analysis that classifies observations in a various cluster through probabilities of belonging to each cluster. Fuzzy clustering algorithm based on the membership degree can produce clusters that better reflect the complex structure of multimorbidity. This study aims to analyse the multimorbidity patterns with fuzzy clustering.

Methods

The study design was observational, retrospective and cross-sectional. The sample was comprised by 525 institutionalised older adults in Spain with very poor physical and mental health. Information about the presence of nine diseases related with dementia, neuropsychiatric disorders, cardiovascular and metabolic diseases was collected: memory problems, Alzheimer, Parkinson, other mental problems, depression, cardiovascular, diabetes, cholesterol and hypertension.

For the statistical analysis, we used the fanny function of cluster package of R. The fanny function is a fuzzy clustering algorithm which allows each observation to spread out over the different cluster and gives a membership probability of this observation, in contrast with other cluster analyses. This algorithm consists in minimizing the following objective function:

$$\sum_{v=1}^{k} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} u_{iv}^{r} u_{jv}^{r} d(i,j)}{2 \sum_{j=1}^{n} u_{jv}^{r}}$$

Where “n” is the number of observations, “k” is the number of clusters and “r” is the membership exponent, which provided crisper clustering when trend to the unity. For these analyses, we used two values of membership, 1.2 and 1.3. Also, “u_{iv}” is the membership of observation i to cluster v, and “d(i,j)” is the dissimilarity between observations i and j. The dissimilarity is calculated through the Jaccard similarity coefficient for binary variables with contingency tables between diseases. We present an example about how this coefficient is calculated in a contingency table between hypertension and cardiovascular problems (Table 1).

Table 1. Contingency table between hypertension and cardiovascular problems

		Hypertension	
		Yes	No
Cardiovascular	Yes	a=163	b=116
	No	c=112	d=129

For this example, the Jaccard similarity coefficient was calculated as follow:

$$s_{ij} = \frac{a}{a+b+c} = \frac{\mathbf{163}}{\mathbf{163+116+112}} \mathbf{= 0.417}$$

Therefore, each value of dissimilarity matrix was:

$$d_{ij} = \sqrt{1 - s_{ij}}$$

Finally, the silhouette of the clusters was analysed. The plot of silhouette is a representation for all i observations descending order (i=1,…,n). A low coefficient of silhouette means the cluster structure is poor, so the clusters are not well differentiated.

Results

The disease with higher prevalence was memory problems (78.1%), followed by mental disease and others related with the dementia (63.1%).

Two clusters were differentiated. The results with membership of 1.2 indicated one cluster with Parkinson's disease, memory problems, depression, Alzheimer, and a second cluster with other mental disorders, cardiovascular, diabetes, cholesterol and hypertension. For membership exponent of 1.3, one cluster was composed by Parkinson's disease, diabetes and cholesterol, and the other cluster by the rest of the diseases.

The average of silhouette was low for the two values of membership exponent, although is higher for membership fuzziness with exponent = 1.3.

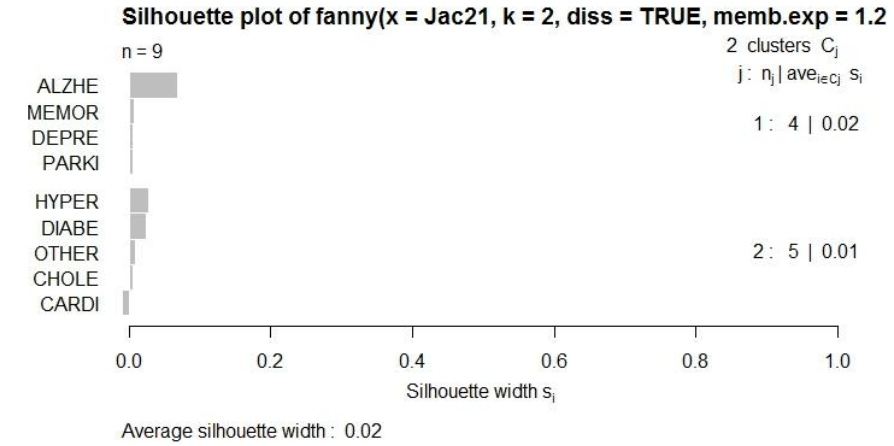

Figure 2a. Silhouette plot of fanny function for institutionalised older adults with dementia (membership exponent=1.2).

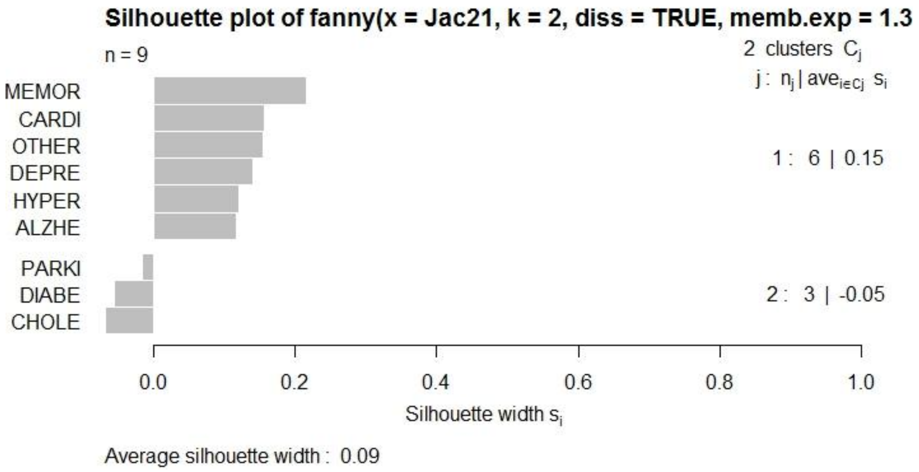

Figure 2b. Silhouette plot of fanny function for institutionalised older adults with dementia (membership exponent=1.3).

Table 3 showed that the membership degree of each cluster is very fuzzy for the value of 1.3. For example, Alzheimer belong to cluster 1 with a probability of 68%, and also to cluster 2 with a 32%. However, this membership is clearer with exponent of 1.2, where the probability of Alzheimer for cluster 1 is 95%.

Table 3. Membership probability of each disease in fuzzy clustering algorithm.

	Membership exponent 1.2		Membership exponent 1.3	
	Cluster 1	Cluster 2	Cluster 1	Cluster 2
Memory	0.92	0.08	0.88	0.12
Depression	0.90	0.10	0.75	0.25
Alzheimer	0.95	0.05	0.68	0.32
Parkinson's disease	0.83	0.17	0.15	0.85
Other Mental disorders	0.12	0.88	0.80	0.20
Cardiovascular	0.14	0.86	0.80	0.20
Diabetes	0.12	0.88	0.16	0.84
Cholesterol	0.15	0.85	0.16	0.84
Hypertension	0.09	0.91	0.70	0.30

Conclusion

This study showed that a fuzzy cluster with higher membership exponent provides better results and the probability of membership was less sharp. The silhouette coefficient is very small in both cases, indicating that there is no structure of patterns of comorbidity. A possible explanation for the lack of the structure could be that there are not groups of diseases that tend to occur jointly in a same patient. One advantage provided by the fuzzy clustering is that the probability of membership could be used in future explanatory models.

Bibliography

1. Jakovljević M. Crnčević Ž (2012) Comorbidity as an epistemological challenge to modern psychiatry. Dialogues Philos. Ment. Neuro Sci.

2. Cornell JE. Pugh JA. Williams Jr JW. et al. (2009) Multimorbidity clusters: clustering binary data from multimorbidity clusters: clustering binary data from a large administrative medical database. Appl Multivar Res 12:163–182.