

1                   **Symptom-based ordinal scale fuzzy clustering of functional gastrointestinal disorders**

2

3 Marjan Mansourian<sup>1,2</sup>, Hamid Reza Marateb<sup>3,4,5,\*</sup>, Ammar Hassanzadeh Keshteli<sup>6</sup>, Hamed Daghigh  
4 Zadeh<sup>7</sup>, Miquel Angel Mañanas<sup>8,9</sup>, Peyman Adibi<sup>10</sup>

5 <sup>1</sup>PhD, Department of Biostatistics and Epidemiology, School of Public Health, Isfahan University of Medical Sciences, HezarJerb St.,  
6 81745 Isfahan, Iran; email: [j\\_mansourian@hlth.mui.ac.ir](mailto:j_mansourian@hlth.mui.ac.ir)

7 <sup>2</sup>Department of Automatic Control, Biomedical Engineering Research Center, Universitat Politècnica de Catalunya BarcelonaTech,  
8 Barcelona, Spain; email: [marjan.mansourian@upc.edu](mailto:marjan.mansourian@upc.edu)

9 <sup>3</sup>PhD, Biomedical Engineering Department, Faculty of Engineering, University of Isfahan, HezarJerb St., 81744 Isfahan, Iran; email:  
10 [h.marateb@eng.ui.ac.ir](mailto:h.marateb@eng.ui.ac.ir).

11 <sup>4</sup>Department of Automatic Control, Biomedical Engineering Research Center, Universitat Politècnica de Catalunya BarcelonaTech,  
12 Barcelona, Spain; email: [hamid.reza.marateb@upc.edu](mailto:hamid.reza.marateb@upc.edu)

13 <sup>5</sup>Department of Machine learning, Isfahan Cardiovascular Research Center, Cardiovascular Research Institute, Isfahan University of  
14 Medical Sciences, Isfahan, Iran.

15 <sup>6</sup>PhD, Department of Medicine, University of Alberta, Edmonton, T6G 2E1 Canada; email: [ahassanz@ualberta.ca](mailto:ahassanz@ualberta.ca).

16 <sup>7</sup>Professor of Medicine, Gastroenterology section, Department of Internal Medicine, Faculty of Medicine, Isfahan University of Medical  
17 Sciences, HezarJerb St., 81745 Isfahan, Iran; email: [daghaghzadeh@mui.ac.ir](mailto:daghaghzadeh@mui.ac.ir).

18 <sup>8</sup>PhD, Department of Automatic Control, Biomedical Engineering Research Center, Universitat Politècnica de Catalunya BarcelonaTech,  
19 Barcelona, Spain; email: [miquel.angel.mananas@upc.edu](mailto:miquel.angel.mananas@upc.edu).

20 <sup>9</sup>Biomedical Research Networking Center in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Barcelona, Spain.

21 <sup>10</sup>Professor of Medicine, Gastroenterology section, Department of Internal Medicine, Faculty of Medicine, Isfahan University of Medical  
22 Sciences, HezarJerb St., 81745 Isfahan, Iran; email: [adibi@med.mui.ac.ir](mailto:adibi@med.mui.ac.ir).

23 \* Address for correspondence:

24                   Hamid Reza Marateb  
25                   Biomedical Engineering Research Center (CREB)  
26                   Building H, Floor 4, Av. Diagonal 647, 08028 Barcelona, Spain  
27                   Phone: +34 658085138  
28                   Email: [hamid.reza.marateb@upc.edu](mailto:hamid.reza.marateb@upc.edu)  
29                   <https://bioart.upc.edu/en/staff/>

30

31

32

33

34

35

## 1 **Abstract**

## 2 **Background**

3           The validity of Rome III criteria for diagnosing functional gastrointestinal disorders  
4 (FGIDs) have been frequently questioned in the literature. In epidemiology, when a disease is  
5 diagnosed, the existence of a true cluster must be proven. Thus, clustering the common GI  
6 symptoms of individuals and comparing the clusters with FGIDs defined by the Rome III  
7 criteria could provide insights about the validity of FGIDs defined by those criteria.  
8 Well-separated compact clusters were detected in responses to questionnaires of the  
9 epidemiological features of different FGIDs in Iranian adults using fuzzy ordinal clustering.  
10 The representative sample from each cluster i.e. Cluster Representative (CR) was formed  
11 whose corresponding FGID was diagnosed with Rome III criteria. Then, FGID diagnosis was  
12 performed for all participants in each cluster and the percentage of cases whose FGID was  
13 the same as the cluster's identified FGID (agreement) was reported.

## 14 **Results**

15           Fourteen valid clusters were detected in 4763 people. The average membership of the  
16 objects in each cluster was 77.3%, indicating similarity of the objects in clusters to their  
17 corresponding CRs. Eight clusters were assigned to single FGIDs (irritable bowel syndromes:  
18 constipation IBS-C, diarrhea IBS-D and un-subtyped IBS-U; functional bloating FB;  
19 functional constipation FC; belching disorder BD. The agreement was higher than 50% in  
20 single FGID clusters except those whose diagnosis was IBS-U.

## 21 **Conclusions**

22           IBS-C, IBS-D, FC, BD, and FB defined with Rome III criteria exist in the population,  
23 which is not the case for IBS-U.

24

1 **Key words:** Cluster analysis; Data mining; Gastrointestinal diseases; Rome III criteria;

2 Validation studies;

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

## 1 **Introduction**

2

3 Throughout documented history, structural diseases could be identified by pathologists  
4 and at times cured by medical technology. However, the nonstructural symptoms described as  
5 “functional” remain enigmatic and less amenable for explanation or effective treatment (1).

6 The research consistency and more reliable diagnosis of FGIDs have been attained  
7 utilizing the Rome criteria in conjunction with self-reported questionnaires (1, 2). The Rome  
8 classification system is based on symptom clusters assumed to remain consistent through  
9 population groups. These criteria have been modified occasionally, and have undergone  
10 many revisions since their first presentation (3).

11

12

13 A number of studies have been performed in the literature on the validity of symptom-  
14 based FGID diagnosis(4-7). Eslick *et al* performed the factor analysis of 46 symptom items  
15 of 897 patients to identify the underlying latent factors. A *k*-means cluster analysis was then  
16 applied on the extracted factors to detect the number of patient groups (4). Alessandro Digesu  
17 *et al* examined whether Rome III criteria are reliable to diagnose constipation in 201 women.  
18 They found out that the Rome III questionnaire did not have significant construct validity  
19 with patients’ self-report of constipation, stool frequency and stool form. They finally  
20 concluded that Rome III is not a valid instrument to diagnose constipation(6). Whitehead and  
21 Drossman performed a critical review on the validity of symptom-based criteria for  
22 diagnosing irritable bowel syndrome (IBS). They concluded that the sensitivity and  
23 specificity of those criteria are 40%-90% and 70%-90%, respectively with clinical diagnoses  
24 made by experienced clinicians (7). Dang *et al* performed a systematic review of symptom-  
25 based diagnostic criteria for IBS. They considered different validation studies. The patient

1 sample size was less than 350 in any studies considered. Based on the manuscripts reviewed  
2 by the authors until 2019, they demonstrated poor validity and utilization of Rome III for  
3 diagnosing IBS(5).

4 Symptom-based diagnosis of distinct FGIDs has been criticized on epidemiology due to  
5 the overlap between symptoms(4). Traditional clustering methods such as *k*-means or  
6 hierarchical algorithms do not allow the overlap between symptoms or FGIDs. This problem  
7 has been solved in data mining by using fuzzy clustering(8). Using advanced Fuzzy data-  
8 mining techniques on large sample sizes, could give new insights on the validity of symptom-  
9 based FGID diagnosis.

10 This study aimed to identify whether FGIDs could be identified based on common  
11 gastrointestinal symptoms of individuals who formed compact and well-separated  
12 epidemiological groups (clusters) in a population.  
13

## 14 **Methods**

15 The Study on the Epidemiology of Psychological, Alimentary Health and Nutrition  
16 (SEPAHAN) project was designed to be a two-phase cross-sectional study. All information  
17 was gathered through self-administered questionnaires mainly aimed to investigate the  
18 epidemiological features of different FGIDs within a large sample of Iranian adults working  
19 in 50 different healthcare centers across Isfahan province, Iran (9). The 4-item rating ordinal-  
20 measurement scale including “never or rarely”, “sometimes”, “often”, and “always” labeled  
21 as 1-4, were used in our analysis.  
22

23 Overall, detailed information from 4763 individuals became available for further  
24 analysis. The subjects did not have any alarm symptoms. All subjects gave informed consent  
25 to the experimental procedure. Ethical approval to conduct the study was provided by the

1 Medical Research Ethics Committee of IUMS (#189069, #189082, and #189086), conformed  
2 to the Declaration of Helsinki.

3 The main research question in our study was as follow: Do FGIDs diagnosed using  
4 Rome III criteria have fuzzy ontology? Also, two research hypotheses were taken into  
5 account in our study. 1) It is possible to accurately determine subject groups based on their  
6 responses to GI questionnaire; and 2) The groups formed by the clustering method could be  
7 assigned to one/more FGIDs, based on the Rome III criteria. Thus, an unsupervised machine  
8 learning approach was developed for the accurate identification of different groups (clusters)  
9 of FGIDs according to SEPAHAN questionnaire items. The particle swarm optimization-  
10 based unsupervised-fuzzy clustering technique was used as the optimum ordinal-to-interval  
11 measurement scale conversion.

12 Data were analyzed using Matlab and Statistics Toolbox Release 2012a (The  
13 MathWorks, Inc., Natick, Massachusetts, USA). Quantitative data were expressed as mean  
14 (SD) and qualitative as number (percent). Determining different FGID clusters according to  
15 SEPAHAN questionnaire using the clustering analysis, 31 significant head questions were  
16 selected by two experts in gastroenterology.

17

18 First, those whose response to *all* of the 31 selected head questions was either “never or  
19 rarely” or ” sometimes” were regarded as “healthy” and excluded from the clustering  
20 analysis. Thus, the following clustering methodology is related to unhealthy subjects. The  
21 matrix (“IN”) was then formed for the unhealthy subjects, in which the labels of the answered  
22 questions for each person were regarded as a row.

23 People having similar symptoms could belong to distinct diagnostic groups (FGID  
24 categories). Fuzzy clustering could be enrolled in which each subject belongs to all of the  
25 FGID groups (clusters) with different membership degrees in [0, 1]. This gives the flexibility

1 to express that objects can belong to more than one cluster. In fact, the terms ‘cluster’ and  
2 ‘cluster membership’ are identical with the epidemiological terms ‘group’ and ‘group  
3 membership’ (4). Also, when a disease is diagnosed in epidemiology, the existence of a true  
4 disease cluster must be proven(10). The clustering algorithm is briefly illustrated as  
5 **Algorithm 1** in the appendices. The clustering method used in our study has been presented  
6 and then validated on ordinal datasets in the literature and showed promising performance by  
7 Brouwer and Groenwold (11) and Marateb *et al* (12).

8 A good partition produces a small value for the compactness, while well-separated  
9 clusters produce a high value for the separation. Thus, the best partition is related to the  
10 minimum of the XB CVI defined as the ratio of compactness to separation (13). After  
11 detecting the best clustering structure, similar clusters were merged using single-linkage  
12 clustering (using the cluster representatives as the input samples)(14). Each subject was then  
13 classified to the cluster for which its membership value was the highest. Like many  
14 epidemiologic studies in which a sample object is required from a target population of  
15 interest(15), we defined the cluster representative of each group as the representative sample  
16 (object) from that cluster. The representative sample was chosen from the objects of  
17 corresponding cluster, where the sum of the distance of the other objects in that cluster to the  
18 representative is minimal. The representative sample was, in fact, the vector of representative  
19 symptoms. The symptoms were then compared with those of ROME III by an expert in  
20 Gastroenterology and the corresponding FGIDs were identified.

## 21 **Results**

22 The socio-demographic and anthropometric factors of the sample population are listed in  
23 Table 1. The percentage of healthy people, as defined in the statistical methods section, was  
24 51.15%. They were excluded from the descriptive and clustering analysis.

25

1 **Table 1:** General characteristics of study participants across different categories of  
2 variables

<b>Variables</b>	<b>categories</b>	<b>Descriptive statistics</b>
<b>Sex</b>	Male	2106(%44.2)
	Female	2657(%55.8)
<b>Marital status</b>	Single	3776(%81.2)
	Married	795(%17.1)
	Divorced or widowed	79 (1.7%)
<b>Educational level</b>	under diploma	638 (13.4%)
	Diploma	1348 (28.3%)
	above diploma but under master	2338 (49.1%)
	master or above	312 (6.6%)
<b>Age (year)</b>	-	36.6±3.1
<b>BMI (kg/m<sup>2</sup>)</b>	-	24.9±5.1
<b>Weight (kg)</b>	-	68.89±13.49

3 All values are mean ± SD for quantitative variables and number (percent for qualitative variables; BMI:  
4 Body mass index;

5 The average membership of the objects in each cluster is shown in Table 2. The  
6 membership value for each object was calculated in terms of the relative similarity (16)  
7 between the 31 ordinal-scale questionnaire answers of that subject and those of the  
8 corresponding cluster representative. It ranged from 0% (complete dissimilarity) to 100%  
9 (complete similarity or clustering consistency).

10 The percentage of people (i.e., the number of people divided by 4763) belonging to  
11 identified clusters is also listed in Table 2. The matched FGIDs identified using the major  
12 symptoms of the subjects in the identified clusters are also shown in Table 2.

1

2 **Table 2:** The identified functional gastrointestinal disorders subtypes and the percentage of people in  
 3 each cluster, and the average membership of the objects in each cluster identified in the SEPAHAN  
 4 dataset shown in mean  $\pm$  SD.

Cluster Index	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
% People <sup>a</sup>	1.46	2.58	1.41	0.61	2.48	0.98	6.12	8.80	2.08	1.14	8.45	8.11	4.41	0.22
Average membership <sup>b</sup> (%)	83 $\pm$ 5	82 $\pm$ 6	72 $\pm$ 6	56 $\pm$ 8	79 $\pm$ 5	77 $\pm$ 5	83 $\pm$ 8	82 $\pm$ 10	77 $\pm$ 5	73 $\pm$ 6	82 $\pm$ 10	83 $\pm$ 7	78 $\pm$ 8	75 $\pm$ 5
FGIDs <sup>c</sup>	IBS-C + FD (non PDS)	IBS- U	IBS-C + FD (PDS)	IBS-C+ FD (PDS)+ UH	IBS- C+ FD (PDS)	IBS- D	FB	FB	IBS- C	IBS-C+ FD (PDS)	BD	IBS- U	FC	IBS- M+ UH

5 <sup>a</sup>: The total percentage of people is 48.9%. They are the “unhealthy” people in the analyzed samples. The rest,  
 6 i.e. “healthy” people were excluded from the analysis as illustrated in the “statistical methods” section. The  
 7 percentage of people was calculated as calculated as the number of people in each cluster divided by the total  
 8 sample size of 4763. <sup>b</sup>: The membership value for each object was calculated in terms of the relative similarity  
 9 between 31 the questionnaire answers of that subject and those of the corresponding cluster representative. <sup>c</sup>:  
 10 Identification of common FGIDs was done by two gastroenterologists. FGIDs: functional gastrointestinal  
 11 disorders, IBS: irritable bowel syndrome, IBS-C: constipation predominant IBS, IBS-D: diarrhea predominant  
 12 IBS, IBS-M: mixed IBS, IBS-U: unsubtyped IBS, FD: functional dyspepsia, FB: functional bloating, FC:  
 13 functional constipation, PDS: postprandial distress syndrome, UH: uninvestigated heartburn, BD: belching  
 14 disorder.

15 Clusters C2, C6, C7, C8, C9, C11, C12, and C13 were assigned to specific FGIDs  
 16 based on the similarity of the cluster representative with Rome III criteria. The agreement  
 17 between their symptoms and those of assigned FGIDs were shown as Venn diagrams in  
 18 Figure 1. In this figure, symptom-based FGID diagnosis was performed for all participants in

1 each cluster based on the Rome III criteria, and the percentage of cases whose FGID was the  
2 same as the cluster's identified FGID was shown. The agreement was higher than the cut-off  
3 similarity level of 50% (17) in all cases except those whose diagnosis was IBS-U (C2, and  
4 C12). C2 and C12 were both assigned to IBS-U, although C12 had minor variations with C2.  
5 For example, stoppage, enema and defecation problems existed in C12 but not C2. In C2  
6 cluster, people often had bloating which was not the case in C12. C7 and C8 were both  
7 assigned to functional bloating. Early satiety, borborygmi and bloating were more frequent in  
8 C8 while dyspepsia, belching, and abdominal/non-cardiac chest pain and fullness were more  
9 frequent in C7. There were insufficient criteria for the diagnosis of functional dyspepsia, IBS  
10 or other FGIDs. They could be considered as functional abdominal bloating based on Rome I  
11 and II criteria, although the name was changed to functional bloating in Rome III criteria  
12 (18).

13

14

**Figure 1 about here**

15

16 Cluster C3 was the combination of IBS-C (52%) and PDS (96%), while clusters C14  
17 included IBS-M (43%) and uninvestigated heartburn (85%), where the agreement percentage  
18 was shown in the parenthesis. Cluster C1 was the combination of IBS-C (7%) and FD (non-  
19 PDS) (34%). On the other hand, cluster C5 was the combination of IBS-C (1%) and FD  
20 (PDS) (74%). Clusters C4 and C10 had the combination IBS-C (24%, 38%), and functional  
21 dyspepsia (PDS) (100%, 85%). However, the symptoms of C10 were not as severe as those  
22 of C4. For example, anal sore, bloating, mucus, defecation problem, enema, incomplete  
23 evacuation, hardness, and abdominal pain were more frequent in C4 in comparison with C10.  
24 In the meanwhile, uninvestigated heartburn (21%) was also present in C4. Two identified

1 clusters C4 and C10, are examples of the concept of severity in IBS introduced by Drossman  
2 *et al*(19).

### 3 **Discussion**

4 The clustering algorithms have been widely used in clinical studies to identify subtypes  
5 of diseases or human disease etiological factors; this is a well-known methodology(20). In  
6 epidemiology, when a disease is first identified, its occurrence as a true cluster (group) must  
7 be proven(10). The validity of the identified clusters was assessed based on 1) XB CVI that  
8 ensures the clusters are compact and well-separated, and 2) the average membership  
9 (similarity) of the objects in each cluster (Table 2; above 77% in single FGID clusters). The  
10 cluster representative was selected similar to the epidemiological studies as the sample  
11 representative of that cluster(15). The existence of the FGIDs with the minimum similarity  
12 level of 50%(17), could be proven in our study. However, since the overlap (similarity) of the  
13 single-FGID clusters with Rome-III defined FGIDs are between 0 %-100 %, their degree of  
14 existence could be stated in the framework of Fuzzy ontology(21). Thus, the main research  
15 question was answered. As a validation of this hypothesis, in total, 21.07% of the population  
16 had IBS (Table 2). A recent patient-based analysis of the same database showed that the  
17 prevalence of IBS was 21.7% using Rome III diagnostic criteria (22).

18 Our results concerning that IBS-C, and D could be symptom-based diagnosed are in  
19 agreement with those of Whitehead *et al* (23) and Whitehead and Drossman (24). However,  
20 this is in contrast with some studies that claimed poor validity and utilization of Rome III  
21 criteria (4, 5, 25). Since the clustering was performed on all of the 31 significant head  
22 questions, high agreement with FGIDs diagnosed by few items in Rome III is very hard to  
23 reach. Thus, like many microbiological studies(17), the cut-off similarity value of 50% was  
24 used in our study. However, tightening the cut-off similarity level to about 70%, which is

1 suitable for distinguishing unambiguous epidemiological disease, the ontology of the  
2 following FGIDs FC, FB and IBS-C could be verified(17).

3 In our study, each subject was finally assigned to a cluster. This is one of the limitations  
4 of our study in which concurrent FGIDs could not be entirely identified. Alternatively, the  
5 membership function could be used to identify different clusters to which a subject belongs  
6 that will be focused in the future.

### 7 **Abbreviations**

8 BD: belching disorder; CR: Cluster Representative; FGIDs: functional gastrointestinal  
9 disorders; FB: functional bloating; FC: functional constipation; IBS: irritable bowel  
10 syndromes; IBS: irritable bowel syndrome, IBS-C: constipation predominant IBS, IBS-D:  
11 diarrhea predominant IBS, IBS-U: unsubtyped IBS; PDS: postprandial distress syndrome,  
12 SEPAHAN: Study on the Epidemiology of Psychological, Alimentary Health and Nutrition;  
13 UH: uninvestigated heartburn;

14

### 15 **Ethics Approval**

16 Written informed consents were obtained from all participants, and the ethics committee of  
17 Isfahan University of medical sciences approved the study.

18

### 19 **Competing Interest**

20 The authors declare that they have no competing interests.

21

### 22 **Funding**

23 The research leading to these results has received funding from the European Union's  
24 Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant  
25 agreement No 712949 (TECNIOspring PLUS) and from the Agency for Business

1 Competitiveness of the Government of Catalonia and also by the Spanish Ministry of  
2 Economy and Competitiveness- Spain project DPI2017-83989-R.

3

#### 4 **Authors Contribution**

5 M.M., HR.M., P.A, and A.H.K designed research; M.M. and HR.M. analyzed data; M.A.M  
6 and M.M. wrote the paper. All authors read and approved the final manuscript.

#### 7 **Acknowledgements**

8 The authors are grateful to Kevin McGill for reviewing a draft of this paper. The  
9 authors are thankful to Sobhan Goudarzi for collaboration in implementing part of the  
10 clustering algorithm.

11

12

#### References

13

- 14 1. Drossman DA. The functional gastrointestinal disorders and the Rome III process.  
15 Gastroenterology. 2006;130(5):1377-90.
- 16 2. Drossman DA, Dumitrascu DL. Rome III: New standard for functional  
17 gastrointestinal disorders. Journal of Gastrointestinal and Liver Diseases. 2006;15(3):237.
- 18 3. Jung HK. Rome III Criteria for Functional Gastrointestinal Disorders: Is There a Need  
19 for a Better Definition? Journal of neurogastroenterology and motility. 2011;17(3):211-2.
- 20 4. Eslick GD, Howell SC, Hammer J, Talley NJ. Empirically derived symptom sub-  
21 groups correspond poorly with diagnostic criteria for functional dyspepsia and irritable bowel  
22 syndrome. A factor and cluster analysis of a patient sample. Alimentary pharmacology &  
23 therapeutics. 2004;19(1):133-40.
- 24 5. Dang J, Ardila-Hani A, Amichai MM, Chua K, Pimentel M. Systematic review of  
25 diagnostic criteria for IBS demonstrates poor validity and utilization of Rome III.  
26 Neurogastroenterology and motility : the official journal of the European Gastrointestinal  
27 Motility Society. 2012;24(9):853-e397.

- 1 6. Digesu GA, Panayi D, Kundi N, Tekkis P, Fernando R, Khullar V. Validity of the  
2 Rome III Criteria in assessing constipation in women. *International urogynecology journal*.  
3 2010;21(10):1185-93.
- 4 7. Whitehead WE, Drossman DA. Validation of symptom-based diagnostic criteria for  
5 irritable bowel syndrome: a critical review. *The American journal of gastroenterology*.  
6 2010;105(4):814-20; quiz 3, 21.
- 7 8. Szczepaniak PS, Lisboa PJ, Kacprzyk J. *Fuzzy systems in medicine*: Springer; 2000.
- 8 9. Adibi P, Keshteli AH, Esmailzadeh A, Afshar H, Roohafza H, Bagherian-Sararoudi  
9 H, et al. The study on the epidemiology of psychological, alimentary health and nutrition  
10 (SEPAHAN): overview of methodology. *J Res Med Sci*. 2012;17(5).
- 11 10. Greenwald P, Block G. Methodological Issues in Epidemiological Studies of Disease  
12 Clusters and Environmental Contamination. In: Vouk VB, Butler GC, Upton AC, Parke DV,  
13 Asher SC, editors. *Methods for assessing the effects of mixtures of chemicals*. Stanford  
14 University: Wiley Chichester; 1987.
- 15 11. Brouwer RK, Groenwold A. Modified fuzzy c-means for ordinal valued attributes  
16 with particle swarm for optimization. *Fuzzy Sets and Systems*. 2010;161(13):1774-89.
- 17 12. Marateb HR, Mansourian M, Adibi P, Farina D. Manipulating measurement scales in  
18 medical statistical analysis and data mining: A review of methodologies. *Journal of Research*  
19 *in Medical Sciences*. 2014;19(1).
- 20 13. Kim D-W, Lee KH, Lee D. On cluster validity index for estimation of the optimal  
21 number of fuzzy clusters. *Pattern Recognition*. 2004;37(10):2009-25.
- 22 14. Duda RO, Hart PE, Stork DG. *Pattern classification*: John Wiley & Sons; 2012.
- 23 15. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*: Lippincott Williams &  
24 Wilkins; 2008.
- 25 16. Ichino M, Yaguchi H. Generalized Minkowski metrics for mixed feature-type data  
26 analysis. *Systems, Man and Cybernetics, IEEE Transactions on*. 1994;24(4):698-708.
- 27 17. Dijkshoorn L, Towner KJ, Struelens M. *New approaches for the generation and*  
28 *analysis of microbial typing data*. Amsterdam ; New York: Elsevier; 2001. xi, 357 p. p.
- 29 18. Seo AY, Kim N, Oh DH. *Abdominal Bloating: Pathophysiology and Treatment*.  
30 *Journal of neurogastroenterology and motility*. 2013;19(4):433-53.
- 31 19. Drossman DA, Chang L, Bellamy N, Gallo-Torres HE, Lembo A, Mearin F, et al.  
32 *Severity in irritable bowel syndrome: a Rome Foundation Working Team report*. *The*  
33 *American journal of gastroenterology*. 2011;106(10):1749-59; quiz 60.

- 1 20. Armstrong RA, Wood L. The identification of pathological subtypes of Alzheimer's  
2 disease using cluster analysis. *Acta neuropathologica*. 1994;88(1):60-6.
- 3 21. Sadegh-Zadeh K. *Handbook of analytic philosophy of medicine*. Dordrecht ; New  
4 York: Springer Verlag; 2012. xxiv, 1133 p. p.
- 5 22. Esmailzadeh A, Keshteli AH, Hajishafiee M, Feizi A, Feinle-Bisset C, Adibi P.  
6 Consumption of spicy foods and the prevalence of irritable bowel syndrome. *World journal*  
7 *of gastroenterology : WJG*. 2013;19(38):6465-71.
- 8 23. Whitehead W, Palsson O, Thiwan S, Talley N, Chey W, Irvine E, et al. Development  
9 and validation of the Rome III diagnostic questionnaire. *Rome III: The Functional*  
10 *Gastrointestinal Disorders 3rd Edition* ed McLean, VA: Degnon Associates, Inc. 2006:835-  
11 53.
- 12 24. Whitehead WE, Drossman DA. Validation of Symptom-Based Diagnostic Criteria for  
13 Irritable Bowel Syndrome: A Critical Review. *Am J Gastroenterol*. 2010;105(4):814-20.
- 14 25. Digesu GA, Panayi D, Kundi N, Tekkis P, Fernando R, Khullar V. Validity of the  
15 Rome III Criteria in assessing constipation in women. *International urogynecology journal*.  
16 2010;21(10):1185-93.  
17  
18  
19  
20

## Appendices

**Algorithm 1:** Identifying the clustering structure of the input matrix data “IN”

For C: =2 to  $\lceil \sqrt{4763} \rceil$ ; where C is the possible number of clusters

    % Identify the clustering structure assuming C clusters as follows:

    % Assign randomly membership degrees to each subject;

    % Initialize OIC matrix “F” using the input data “IN”;

    % Repeat until the algorithm has converged i.e. membership degrees do not change in two consecutive iterations (max. iterations: =10);

        % Estimate “F” matrix so as to minimize the FCM objective function using

        PSO;

        % Update the membership degrees and clusters’ prototype;

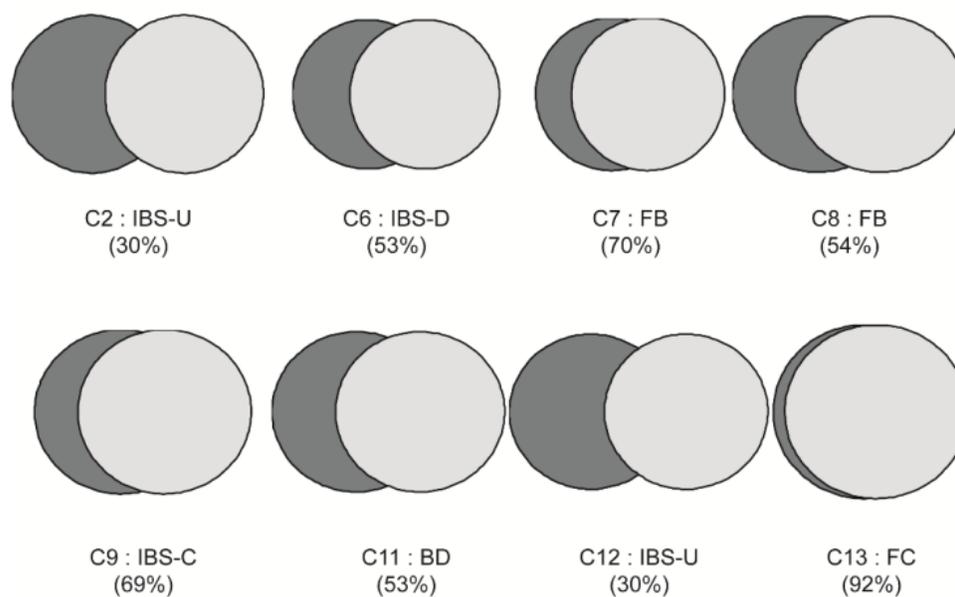
    END Repeat;

    % Calculate the ratio between the compactness of the fuzzy partition and the strength of the separation between clusters (XB Clustering-Validity Index (CVI));

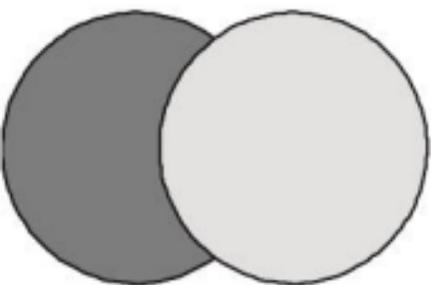
END For;

% Select the optimum number of clusters ( $C_{optm}$ ) based on the minimum value of the XB CVI and save the corresponding clustering structure (“F” matrix, subject’s membership degrees and clusters’ prototypes);

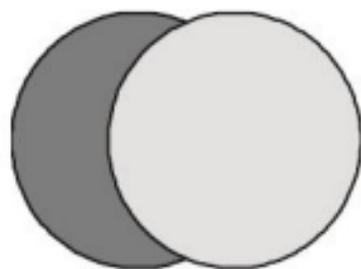
**Figure 1:** The overlap between identified clusters and corresponding functional gastrointestinal disorders (FGIDs) (Based on the Rome III criteria) shown as a Venn diagram. The percentage of the cases whose symptom-based diagnosis was identical to those of their cluster representatives were reported for clusters assigned to a single FGID. IBS: irritable bowel syndrome, IBS-C: constipation predominant IBS, IBS-D: diarrhea predominant IBS, IBS-U: unsubtyped IBS, FB: functional bloating, FC: functional constipation, PDS: postprandial distress syndrome, UH: uninvestigated heartburn, BD: belching disorder.



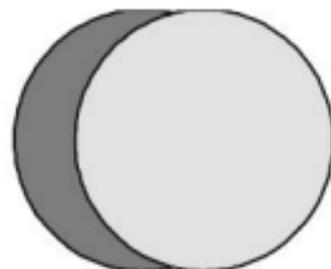
**Figure 1**



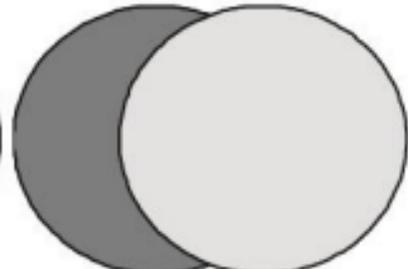
**C2 : IBS-U**  
(30%)



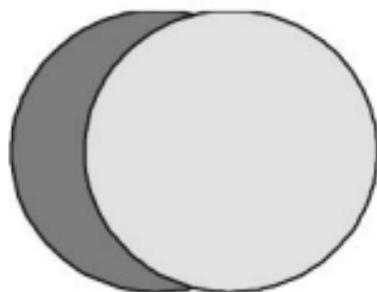
**C6 : IBS-D**  
(53%)



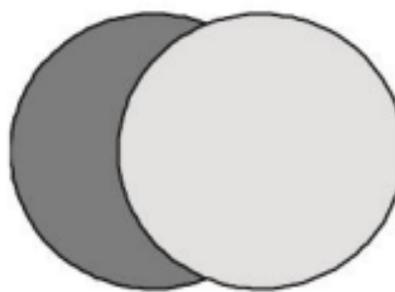
**C7 : FB**  
(70%)



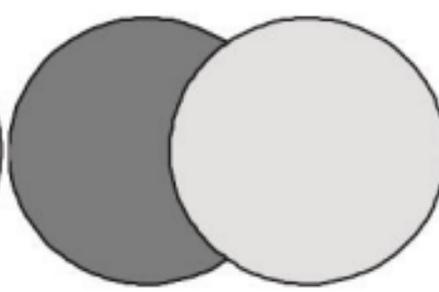
**C8 : FB**  
(54%)



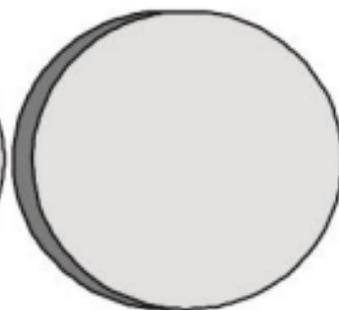
**C9 : IBS-C**  
(69%)



**C11 : BD**  
(53%)



**C12 : IBS-U**  
(30%)



**C13 : FC**  
(92%)