THE PRIMARY CARE COMPANION
FOR CNS DISORDERS

Supplementary Material

Article Title: Phenotyping of Review-Of-Systems Responses to Differentiate Functional Seizures From Epilepsy

Author(s): Wesley T. Kerr, MD, PhD; Corinne H. Allas, BS; Siddhika S. Sreenivasan, BS; Emily A. Janio, BS; Amir H. Karimi, BS; Ishita Dubey, MS; Chelsea T. Braesch, PA-C; Justine M. Le, PA-C; Jessica M. Hori, RN; Akash B. Patel, DO; Norma L. Gallardo, BS; Janar Bauirjan, BSN, RN; Jerome Engel, Jr, MD, PhD; John M. Stern, MD; and Jamie D. Feusner, MD

DOI Number: https://doi.org/10.4088/PCC.21m03049

List of Supplementary Material for the article

1. Supplementary Figure 1
2. Supplementary Figure 2
3. Supplementary Table 1

Disclaimer

This Supplementary Material has been provided by the author(s) as an enhancement to the published article. It has been approved by peer review; however, it has undergone neither editing nor formatting by in-house editorial staff. The material is presented in the manner supplied by the author.

Supplementary Figure 1

UCLA Medical Center

Neurological Services

REVIEW OF SYSTEMS

Please place a checkmark if you currently have any of the following symptoms. (Disregard the bold headings in quotes on the left. They are for administrative purposes only) :

1. "constitutional"	_____ fever	_____ weight loss	_____ fatigue
2. "eyes problems"	_____ blurred vision	_____ double vision	_____ loss of vision
	_____ eye pain	_____ eye redness	_____ eye dryness
3. "ear/nose/throat"	_____ trouble hearing	_____ ringing in ear(s)	_____ dizziness (vertigo)
	_____ loss of balance	_____ ear pain	_____ ear discharge
	_____ hoarseness	_____ trouble swallowing	_____ slurred speech
4. "cardiovascular"	_____ chest pain	_____ irregular heart beat	_____ fast heart beat
	_____ limb swelling	_____ limb pain on walking	_____ fainting
5. "respiratory"	_____ trouble breathing	_____ chronic cough	_____ coughing blood
6. "gastrointestinal"	_____ indigestion	_____ heart burn	_____ abdominal pain
	_____ nausea	_____ vomiting	_____ regurgitation
	_____ diarrhea	_____ constipation	_____ bloody stools
7. "genitourinary"	_____ incontinence	_____ pain on urination	_____ blood in urine
8. "musculoskeletal"	_____ muscle pain	_____ muscle cramp	_____ muscle twitches
	_____ loss of muscle bulk	_____ neck pain	_____ back pain
	_____ joint pain	_____ joint stiffness	_____ joint swelling
9. "skin & breast"	_____ numbness	_____ tingling	_____ discoloration
	_____ hair loss	_____ nail changes	_____ sweating changes
10. "neurologic"	_____ headache	_____ face pain	_____ face numbness
	_____ weakness	_____ tremors	_____ clumsiness
	_____ blackouts	_____ trouble with memory	_____ trouble concentrating
11. "psychiatric"	_____ hallucinations	_____ feeling depressed	_____ trouble sleeping
	_____ suicidal thoughts	_____ inappropriate crying	_____ inappropriate laughing
12. "hematologic/lymphatic"	_____ abnormal bleeding	_____ nose bleeds	_____ lumps or swellings
13. "allergic/immunologic"	_____ skin rash	_____ joint pain	_____ dry eyes &/or dry mouth
14. "endocrine"	_____ excessive thirst	_____ heat or cold intolerance	_____ excessive urination

*Person completing questionnarie*_____ *Relationship to patient*_____

For office use: This questionnaire may be completed by the patient, relatives or ancillary staff provided that it is signed an dated by the treating physician. (Reference may later be made to this information by a signed and dated statement by the treating physician, designating location of the information, date obtained and any subsequent changes.)

*Physician's Signature*_____ *Date*_____

3-1574 MDR (10/91) S
MR C/009 (B)

09/08

Supplementary Figure 2. Methods

In KM, the number of clusters, *k*, is pre-selected and the algorithm is initiated by *k*-data points from *n*-patients being randomly selected as initial medoids [1]. The percent of responses that differs from that medoid is called the Hamming distance [2]. The other (*n*-*k*) patients are assigned to each medoid based on the minimum Hamming distance. With these assignments, a new set of *k*-medoids are set based on the median response within each cluster. K-medoids is preferred over k-means for categorical data. The algorithm of assignment to clusters is repeated iteratively until convergence. We evaluated KM for pre-defined *k* ranging from 2 to 70. The upper limit of 70 was chosen to facilitate statistical identifiability of representing 78 ROS symptoms into a slightly smaller number of clusters. Validation data was assigned to the closest medoid using a Hamming distance.

In LCA, the number of latent classes, *k*, also is pre-selected and the algorithm is initiated by data points being randomly assigned to *k*-clusters [3]. Cluster membership was determined based on a sequential expectation-maximization (EM) algorithm. In the expectation (E)-step, the expected cluster sizes and assignments are calculated conditional on the definitions of how each symptom probabilistically contributes to the likelihood that each patient lies within each cluster. In the maximization (M)-step, those definitions are modified to maximize the likelihood of the expected cluster assignments determined in the E-step. The E and M steps are applied sequentially until convergence in the MDLV toolbox [4]. Within this convergence method, cluster membership is probabilistic, whereas after convergence, patients were assigned greedily the cluster with highest probability of membership. We varied *k* from 2 to 70, similar to KM. Validation data was assigned to the cluster with the highest membership probability.

Supplementary Table 1. Exact performance of each method evaluated using leave-one-out cross-validation. Confidence intervals estimated with binomial exact statistics or, for the area under the receiver operating curve (AUC), the Wald method. Abbreviations: Logistic regression based on individual symptoms (LR), LR with recursive feature elimination (LR-RFE), k-medoids (KM), principal component analysis (PCA), independent component analysis (ICA), latent class analysis (LCA), density-based spatial clustering of applications with noise (DBSCAN), estimate (Est), confidence interval (CI), predictive value (PV).

	Accuracy		Sensitivity		Specificity		ES-PV		DS-PV		AUC	
Method	Est (%)	95% CI	Est (%)	95% CI	Est (%)	95% CI	Est (%)	95% CI	Est (%)	95% CI	Est (%)	95% CI
LR	76	72-81	88	84-92	35	25-45	82	78-87	46	34-58	62	53-69
LR-RFE	75	71-80	91	87-94	24	15-33	80	76-85	42	28-56	64	55-72
Total	79	75-84	97	95-99	17	9-25	80	75-85	64	44-84	72	65-78
KM	78	74-83	97	95-100	13	6-21	79	75-84	60	40-80	69	61-76
PCA	79	74-83	85	81-90	55	45-66	87	82-91	53	42-63	74	66-80
ICA	47	41-52	35	30-41	87	79-94	90	84-96	28	22-34	67	60-74
LCA	80	76-85	96	93-98	28	19-39	82	77-86	65	50-80	72	65-79
DBSCAN	77	67-87	100	100-100	0	0-0	77	73-82	-	-	69	61-76

In DBSCAN, we eschew the concept of pre-defining the number of clusters and summarizing a cluster by a central point or pattern of responses [5]. Instead, we pre-define an epsilon, ε, and a minimum number of points, *m*, needed to define a cluster. We randomly select data points and assign any data points within a Hamming distance of ε to the same cluster. We extend the definition of that cluster by iteratively extending the cluster to include all data points within ε of the points within the presumptive cluster. This iterative process stops when there are no more points within ε of the points within the presumptive cluster and the cluster is maintained if it includes at least *m* points. If the cluster is too small, the data points are assigned to the "outlier" cluster. This process of cluster determination is repeated for each data point. We vary ε from the minimum non-zero to maximum observed Hamming distance in the dataset. We vary *m* from 2 to the entire size of the entire dataset, minus one point. If validation data was within ε of data within a cluster, then it was assigned to that cluster but otherwise it was considered an "outlier."

In PCA and ICA, the hypothesis is that the 78 ROS symptoms can be summarized using a lower number of components based on patterns in similar responses across multiple symptoms. Components are interpreted as combinations of similar symptoms that may be more interpretable than individual symptoms (e.g. a pain component including musculoskeletal, joint, and head pain). In PCA, the first component is determined based on the single vector that maximizes the variance of the data when projected onto that vector [6]. Each subsequent component is determined based on the vector perpendicular to all prior components that maximizes the remaining variance of the projected data. In ICA, the components are initialized with the PCA components [7]. These components are modified

iteratively by the FAST ICA algorithm to maximize the statistical independence of components [7], as compared to the requirement of being perpendicular. This tends to lead to a sparse representation of the data. Validation data was projected onto these learned components and did not contribute to the determination of components.

References from Supplemental Methods:

[1] Kaufman L, Rousseeuw PJ. Clustering by Means of Medoids. In: Dodge Y, editor. Statistical Data Analysis Based on the L1-Norm and Related Methods. North-Holland; 1987, p. 405-416.

[2] Hamming RW. Error detecting and error correcting codes. The Bell System Technical Journal 1950;29: 147-160.

[3] Lazarsfeld PF, Henry NW. Latent Structure Analysis. New York: Houghton Mifflin; 1968.

[4] Yu HT. Models with discrete latent variables for analysis of categorical data: a framework and a MATLAB MDLV toolbox. Behav Res Methods 2013;45: 1036-47.

[5] Ester M, Kriegel H-P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Simoudis E, Han J, Fayyad U, editors. Second International Conference on Knowledge Discovery and Data Mining (KDD-96): AAAI Press; 1996. p. 226-231.

[6] Pearson K. On lines and planes of closest fit to systems of points in space. Philosophical Magazine 1901;2: 559-572.

[7] Hyvarinen A. Fast and robust fixed-point algorithms for independent component analysis. Ieee Transactions on Neural Networks 1999;10: 626-634.