MINISTERUL EDUCAŢIEI NAŢIONALE
ROMÂNIA
UNIVERSITATEA DE MEDICINĂ, FARMACIE, ŞTIINŢE ŞI TEHNOLOGIE „GEORGE EMIL PALADE" DIN TÂRGU MUREŞ

Acta Marisiensis. SeriaTechnologica
Vol. 18 (XXXV) no. 1, 2021
ISSN 2668-4217, ISSN-L 2668-4217

sciendo

10.2478/amset-2021-0008

EXPLORATORY FACTOR ANALYSIS FOR IDENTIFYING COMORBIDITIES AS RISK FACTORS AMONG PATIENTS WITH CIED

Bernadett-Miriam DOBAI[1], Laszlo Barna IANTOVICS[2*], Andreea PAIU[3]

[1,2] *"George Emil Palade" University of Medicine, Pharmacy, Science and Technology of Targu Mures*

Gheorghe Marinescu street, 38, Targu Mures, Romania

[1]dobai.bernadettmiriam@gmail.com

[2*] Corresponding author, barna_iantovics@yahoo.com

[3] *"Transilvania" University of Brasov*

B-dul Eroilor, 29, Brasov, Romania

[3]andreea.paiu@gmail.com

Abstract

The emergence of SARS-CoV-2 affected care both for acute and chronic health conditions. Majority of the patients with cardiac implantable electronic devices (CIEDs) have multiple comorbidities, which can influence their response to COVID-19. An online survey consisting of 45 multiple-choice question was designed for CIED patients assessing comorbidities and overall health condition during September -December 2020. A multivariate analysis based on principal axis factoring (PAF) was performed on the eligible 184 survey response. Three factors were identified. Ten-year survival rates were calculated with Charlson Comorbidity Index. The extracted factors explained 66.1% of the cumulative variance and were consistent with medical literature data.

Key words: SARS-CoV-2, COVID-19, CIED, survival rate, exploratory factor analysis, principal axis factoring, Industry 4.0, smart manufacturing

1. Introduction

The recent emergence of SARS-CoV-2 affected quality of care for both acute and chronic health conditions. Cardiac implantable electronic devices (CIED) are the gold standard therapy for various cardiac conditions, implantation rates showing a year-by-year increase worldwide [1, 2]. Majority of the patients with CIEDs have multiple underlying health conditions (comorbidities) which are considered as risk factors. These might increase the risk of severe illness from COVID-19 [3]. Therefore, a survey consisting of 45 multiple-choice questions was designed to evaluate CIED patients' comorbidities and overall health status during COVID-19 pandemic. According to the Charlson Comorbidity index [4-6] respondents' 10-year survival rate was calculated, based on the presence of their underlying health conditions and age.

In the present paper, a multivariate analysis based on a principal axis factoring (PAF) was performed regarding CIED patients' comorbidities, on the following variables: arterial hypertension (AHT), diabetes mellitus (DM), obesity, heart failure (HF), chronic obstructive pulmonary disease (COPD), smoking, malignancy. PAF belongs to the class of Exploratory Factor Analysis (EFA). EFA methods are frequently used in healthcare research [7-10].

One of the most difficult problems in EFA is the selection of the optimal number of factors [11-15]. Identification of an inappropriate number of factors may lead to inaccurate conclusions. A previous article studied the importance of the total cumulative variance that should be explained by the selected most appropriate number of extracted factors [16]. It was considered that, at least, a minimum threshold of cumulative variance should be explained by the extracted factors that depends on the specificity of the research. Based on this previous study and on a thorough literature documentation, we proposed in the current article three rules to be considered in selecting the most appropriate number of factors.

2. Material and methods

A survey consisting of 45 multiple choice questions was completed online, anonymously and voluntarily by CIEDs patients from the outpatient care of the Institute of Cardiovascular Emergencies and Transplant (IUBCvT) of Targu Mures and by international CIED patients, part of online support groups between September-December 2020. The following type of data was collected: demographic data, health condition and psychosocial impact of COVID-19 pandemic.

A Principal axis factoring (PAF) was performed, applying the varimax [17] orthogonal rotation method focusing on the comorbidities of the respondents. Initially, the study design included 7 variables (Var), that were denoted as: Var1 - AHT; Var2 - DM; Var3 - Obesity; Var4 - COPD; Var5 - HF; Var6 - Smoking; Var7 - Malignancy.

Ten-year (y) survival rates were calculated with Charlson comorbidity index (CCI), used to calculate 10-year survival rates based on respondents' comorbidities and age. The CCI considers several variable, such as age, myocardial infarction, HF, peripheral vascular disease, cerebrovascular accident, dementia, COPD, connective tissue disease, peptic ulcer, liver disease, DM, hemiplegia, kidney disease, malignancies, acquired immunodeficiency syndrome (AIDS) [4, 5]. Although we questioned the presence of multiple risk factors of severe COVID-19 illness, such as AHT, DM, Obesity, COPD, HF, Smoking, Malignancy; the CCI included 3 from these, namely DM, HF, COPD. Based on these variables CCI was calculated.

The assessed number of responses in the data analysis was 184, although the initial number of participants in the survey was 210; 26 participants were excluded due to contradictory responses. In a preliminary analysis Var7 after extraction had a low communality (0.031), consequently it was removed.

Based on the in-depth bibliographic study of the scientific literature, and a previous research [16], we took into account the following three rules to determine the number of extracted factors:

-Rule 1): the extracted eigenvalues to be at least 1, called Kaiser criterion [17].

-Rule 2): visual interpretation of the Scree plot, called Cattell's Scree test [18].

-Rule 3): the total variance explained to be at least 60%-65% for the current study also since none of the variables passed the normality assumption.

3. Results and Discussion

The eligible 184 responses were registered from different regions around the world, mainly from Europe (Romania, Hungary, UK, Ireland, Switzerland, Slovakia, Spain) and the USA; all respondents were patients living with cardiac implantable electronic devices.

The results of the PAF analysis indicated three factors, which explained 66.1% of the cumulative variance, corresponding to rule 3.

Table 1 presents the correlation coefficients (r) between all the variables, and the corresponding significance (sig) of each correlation. We consider the value of significance level as 0.05, sig < 0.05 indicates a significant result. The calculated determinant (det) of the correlation matrix is 0.667.

Figure 1 presents the Scree plot; this line plot shows the eigenvalues of factors and it is used to determine the number of factors to retain in PAF. This visual interpretation (Rule 2) facilitates factor choosing process. According to rule 1, eigenvalues at least 1 or above 1 should be chosen.

Table 2 presents the result of the Bartlett's Test of Sphericity. The obtained small value (~0) of the significance level of the Bartlett's test indicates that the factor analysis can be applied on the considered data. At the same time, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) test was applied. The obtained result was 0.591, which also indicates the correctness of the factor analysis application process.

Table 3 presents the communalities: initially and after extraction. According to the established rules three factors were extracted.

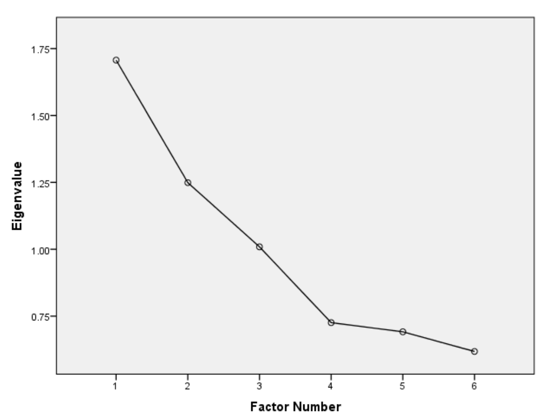

Fig. 1: Scree plot of the eigenvalues

Table 1: Correlation coefficients between all variables

	Var1 (AHT)	Var2 (DM)	Var3 (Obesity)	Var4 (COPD)	Var5 (HF)	Var6 (Smoking)
Var1 (AHT) r (sig)	1 (-)	0.252 (0)	0.073 (0.161)	0.238 (0.001)	0.164 (0.013)	0.233 (0.001)
Var2 (DM) r (sig)	0.252 (0)	1 (-)	0.067 (0.182)	0.224 (0.001)	-0.013 (0.431)	0.176 (0.008)
Var3 (Obesity) r (sig)	0.073 (0.161)	0.067 (0.182)	1 (-)	0.326 (0)	0.045 (0.271)	-0.075 (0.156)
Var4 (COPD) r (sig)	0.238 (0.001)	0.224 (0.001)	0.326 (0)	1 (-)	0.059 (0.214)	0.024 (0.375)
Var5 (HF) r (sig)	0.164 (0.013)	-0.013 (0.431)	0.045 (0.271)	0.059 (0.214)	1 (-)	0.170 (0.011)
Var6 (Smoking) r (sig)	0.233 (0.001)	0.176 (0.008)	-0.075 (0.156)	0.024 (0.375)	0.170 (0.011)	1 (-)

Table 2: Bartlett's Test of Sphericity

Approx. Chi-Square	72.94
df	15
Sig	0

Table 3: The obtained communalities

	Initial	Extraction
Var1 (AHT)	0.151	0.302
Var2 (DM)	0.114	0.317
Var 3 (Obesity)	0.115	0.278
Var4 (COPD)	0.177	0.465
Var5 (HF)	0.054	0.296
Var6 (Smoking)	0.100	0.268

Table 4 presents the extracted eigenvalues and the total cumulative variance explained.

Table 4: presents the total variance explained (selected factors marked with *)

Factor	Initial Eigenvalues		
	Total	%of Variance	Cumulative %
Factor 1*	1.707	28.448	28.448
Factor 2*	1.249	20.815	49.263
Factor 3*	1.009	16.816	66.079
Factor 4	0.726	12.095	78.175
Factor 5	0.692	11.525	89.700
Factor 6	0.618	10.300	100.000

A Rotated Factor Matrix (RFM) table includes the rotated factor loadings that have the significance of correlation strength between the variables and the factors. Values ranged between [-1, +1]. We considered only the correlations with an absolute value above 0.2. Very low correlations were not considered. Table 5 presents the rotated factor matrix and includes the variables loading to the three factors.

Table 5: Rotated factor matrix

	Factor 1	Factor 2	Factor 3
Var2(DM)	0.538		
Var1(AHT)	0.466		0.243
Var6 (Smoking)	0.405		0.288
Var4(COPD)	0.265	0.628	
Var 3(Obesity)		0.526	
Var5(HF)			0.539

Var2 (DM), Var1 (AHT), Var6 (Smoking), and Var4 (COPD) were loaded in Factor 1. These 4 comorbidities are frequently found together. Factor 1 is consistent with medical literature, COPD is associated with DM and AHT, smoking being a well-known major risk factor [19, 20].

Var4 (COPD) and Var 3 (Obesity) were loaded in Factor 2. The correlation between COPD and obesity is researched frequently; an interaction between abnormal adipose tissue function and COPD [21, 22] is hypothesized. The components of Factor 2 justify the correlation between the two diseases.

Var5 (HF), Var6 (Smoking), and Var1 (AHT) were loaded in Factor 3. Smoking is a risk factor for cardiovascular diseases, according to studies, it leads to systolic dysfunction, aggravating HF. The components of Factor 3 emphasize the correlation between the comorbidities [23, 24].

Table 6 shows survival rates calculated for each factor depending on age, which highlights the importance of patients' age on prognosis [4]. The three factors and its' variables depicting different survival rates were consistent with literature data. The riskiest comorbidity is cardiovascular disease (HF), followed by COPD; other risk factors are AHT, DM, obesity.

Table 6: Survival rates calculated with CCI in different age groups

	Factor 1	Factor 2	Factor 3
<50	77%	96%	96%
50-59 year	53%	90%	90%
60-69 year	21%	77%	77%
70-79 year	2%	53%	53%
>80 year	0%	21%	21%

4. Conclusions

The current paper described a multivariate analysis based on Principal Axis Factoring performed on CIEDs patients' comorbidity data. PAF belongs to Exploratory Factor Analysis that is frequently used in healthcare research.

Although, one of the difficulties in PAF is the selection of the number of extracted factors, we proposed three rules to be considered: 1) the extracted eigenvalues to be the value at least 1; 2) the visual interpretation of the scree plot; 3) the cumulative variance explained by the selected factors to be at least 65% (for our research, based on the fact that none of the variables passed the normality assumption).

Three factors were identified. Ten-year survival rates were calculated according to CCI. The extracted factors explained 66.1% of the cumulative variance. The factors' structure was consistent with multiple other studies stating and emphasizing these variables as risk factors.

In this article an experimental study was performed on the data obtained via the questionnaire. The research methodology was based on statistical modeling using exploratory factor analysis. In a forthcoming research, the statistical modeling applied in the current paper together with other methods of artificial intelligence will be applied for industrial big data analysis in the framework of the SOON (Social Network of Machines) project, which is focused on investigating innovative solutions based on the use of intelligent agents with social abilities to optimize the manufacturing processes in smart factories.

Acknowledgement

This work was developed in the framework of the CHIST-ERA program supported by the Future and Emerging Technologies (FET) program of the European Union through the ERA-NET Cofund funding scheme under the grant agreements, title: Social Network of Machines (SOON). This research was supported by a grant of the Romanian National Authority for Scientific Research and Innovation CCDI-UEFISCDI, project number 101/2019, COFUND-CHISTERA-SOON, within PNCDI III.

We would like to thank the "Advanced Research in Information Technology" Research Center (CARIT) and the Electrophysiology Team of the Emergency Institute for Cardiovascular Disease and Transplant (IUBCvT) Targu Mures for the support.

References

[1] Priori, S.G., Blomström-Lundqvist, C., Mazzanti, A., Blom, N., Borggrefe, M., Camm, J., Elliott, P.M., Fitzsimons, D., Hatala, R., Hindricks, G., Kirchhof, P., Kjeldsen, K., Kuck, K.H., Hernandez-Madrid, A., Nikolaou, N., Norekvål, T.M., Spaulding, C. and Van Veldhuisen, D.J. (2015), 2015 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: The Task Force for the Management of Patients with Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death of the European Society of Cardiology (ESC). Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), *Eur Heart J*, vol. 36(41), pp. 2793-2867.

[2] Members, A.T.F., Brignole, M., Auricchio, A., Baron-Esquivias, G., Bordachar, P., Boriani, G., Breithardt, O.-A., Cleland, J., Deharo, J.-C., Delgado, V., Elliott, P.M., Gorenek, B., Israel, C.W., Leclercq, C., Linde, C., et al. (2013), 2013 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy: The Task Force on cardiac pacing and resynchronization therapy of the European Society of Cardiology (ESC). Developed in collaboration with the European Heart Rhythm Association (EHRA), *European Heart Journal*, vol. 34(29), pp. 2281-2329.

[3] Guan, W.J., Liang, W.H., Zhao, Y., Liang, H.R., Chen, Z.S., Li, Y.M., Liu, X.Q., Chen, R.C., Tang, C.L., Wang, T., Ou, C.Q., Li, L., Chen, P.Y., Sang, L., Wang, W., et al. (2020), Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis, *Eur Respir J*, vol. 55(5), pp. 2000547.

[4] Charlson, M.E., Pompei, P., Ales, K.L. and MacKenzie, C.R. (1987), A new method of classifying prognostic comorbidity in longitudinal studies: development and validation, *J Chronic Dis*, vol. 40(5), pp. 373-383.

[5] Quan, H., Li, B., Couris, C.M., Fushimi, K., Graham, P., Hider, P., Januel, J.M. and Sundararajan, V. (2011), Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries, *Am J Epidemiol*, vol. 173(6), pp. 676-682.

[6] Emami, A., Javanmardi, F., Pirbonyeh, N. and Akbari, A. (2020), Prevalence of Underlying Diseases in Hospitalized Patients with COVID-19: a Systematic Review and Meta-Analysis, *Arch Acad Emerg Med*, vol. 8(1), pp.e35.

[7] Angst, J., Adolfsson, R., Benazzi, F., Gamma, A., Hantouche, E., Meyer, T.D., Skeppar, P., Vieta,

E. and Scott, J. (2005), The HCL-32: towards a self-assessment tool for hypomanic symptoms in outpatients, *Journal of affective disorders*, vol. 88(2), pp. 217-233.

[8] An, D., Hong, K.S. and Kim, J.-H. (2011), Exploratory factor analysis and confirmatory factor analysis of the korean version of hypomania checklist-32, *Psychiatry investigation*, vol. 8(4), pp. 334-339.

[9] Prather, J.C., Lobach, D.F., Goodwin, L.K., Hales, J.W., Hage, M.L. and Hammond, W.E. (1997), Medical data mining: knowledge discovery in a clinical data warehouse, *Proceedings : a conference of the American Medical Informatics Association. AMIA Fall Symposium*, pp. 101-105.

[10] Gil-Gouveia, R., Parreira, E. and Martins, I. (2005), Autonomic features in cluster headache. Exploratory factor analysis, *The journal of headache and pain*, vol. 6, pp. 20-23.

[11] Yunjin, C., Jonathan, T. and Robert, T. (2017), Selecting the number of principal components: Estimation of the true rank of a noisy matrix, *The Annals of Statistics*, vol. 45(6), pp. 2590-2617.

[12] Costello, A.B. and Osborne, J. (2005), Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis, *Practical Assessment, Research & Evaluation*, vol. 10, pp. 1-9.

[13] Hayton, J.C., Allen, D.G. and Scarpello, V. (2004), Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis, *Organizational Research Methods*, vol. 7(2), pp. 191-205.

[14] Preacher, K.J., Zhang, G., Kim, C. and Mels, G. (2013), Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective, *Multivariate Behavioral Research*, vol. 48(1), pp. 28-56.

[15] Song, J. and Belin, T.R. (2008), Choosing an appropriate number of factors in factor analysis with incomplete data, *J Comput. Stat. Data Anal.*, vol. 52(7), pp. 3560-3569.

[16] Iantovics, L.B., Rotar, C. and Morar, F. (2019), Survey on establishing the optimal number of factors in exploratory factor analysis applied to data mining, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9(2), pp. e1294.

[17] Kaiser, H.F. (1958), The varimax criterion for analytic rotation in factor analysis, *Psychometrika*, vol. 23(3), pp. 187-200.

[18] Cattell, R.B. (1966), The Scree Test For The Number Of Factors, *Multivariate Behavioral Research*, vol. 1(2), pp. 245-276.

[19] Mahishale, V., Angadi, N., Metgudmath, V., Eti, A., Lolly, M. and Khan, S. (2015), Prevalence and impact of diabetes, hypertension, and cardiovascular diseases in chronic obstructive pulmonary diseases: A hospital-based cross-section study, *J Transl Int Med*, vol. 3(4), pp. 155-160.

[20] Gayle, A., Dickinson, S., Poole, C., Pang, M., Fauconnot, O. and Quint, J.K. (2019), Incidence of type II diabetes in chronic obstructive pulmonary disease: a nested case-control study, *NPJ Prim Care Respir Med*, vol. 29(1), pp. 28.

[21] Franssen, F.M.E., O'Donnell, D.E., Goossens, G.H., Blaak, E.E. and Schols, A.M.W.J. (2008), Obesity and the lung: 5 · Obesity and COPD, vol. 63(12), pp. 1110-1117.

[22] Zewari, S., Vos, P., van den Elshout, F., Dekhuijzen, R. and Heijdra, Y. (2017), Obesity in COPD: Revealed and Unrevealed Issues, *Copd*, vol. 14(6), pp. 663-673.

[23] Kamimura, D., Cain, L.R., Mentz, R.J., White, W.B., Blaha, M.J., DeFilippis, A.P., Fox, E.R., Rodriguez, C.J., Keith, R.J., Benjamin, E.J., Butler, J., Bhatnagar, A., Robertson, R.M., Winniford, M.D., Correa, A. and Hall, M.E. (2018), Cigarette Smoking and Incident Heart Failure: Insights From the Jackson Heart Study, *Circulation*, vol. 137(24), pp. 2572-2582.

[24] Fagard, R.H. (2009), Smoking amplifies cardiovascular risk in patients with hypertension and diabetes, *Diabetes Care*, vol. 32 Suppl 2(Suppl 2), pp. S429-431.