



DOI: 10.4274/gulhane.galenos.2023.69926
Gulhane Med J 2024;66:62-67

Latent profile analysis for the classification of OECD countries with health indicators

© Hülya Özen¹, © Dođukan Özen²

¹University of Health Sciences Türkiye, Gülhane Faculty of Medicine, Department of Medical Informatics, Ankara, Türkiye

²Ankara University Faculty of Veterinary Medicine, Department of Biostatistics, Ankara, Türkiye

Date submitted:

11.09.2023

Date accepted:

02.11.2023

Online publication date:

15.06.2024

Corresponding Author:

Hülya Özen, PhD, University of Health Sciences Türkiye, Gülhane Faculty of Medicine, Department of Medical Informatics, Ankara, Türkiye
+90 312 567 15 46
hulya.ozen@sbu.edu.tr

ORCID:

orcid.org/0000-0003-4144-3732

Keywords: Classification, health equipment, healthcare workers, latent profile analysis, OECD

ABSTRACT

Aims: Health indicators provide up-to-date information on the health status of a population. This study aimed to classify the Organization for Economic Co-operation and Development (OECD) countries according to health indicators and assess their status.

Methods: The dataset was obtained from the OECD and World Bank databases. The most recent data from 2018 to 2022 were used. The dataset included the number of hospital beds, computed tomography scanners, magnetic resonance imaging (MRI) units, mammography machines, and radiotherapy machines as indicators of health equipment and the number of doctors, nurses, medical graduates, and nursing graduates as indicators of healthcare workers. The classification was performed using latent profile analysis (LPA). Estimated classes were compared using ANOVA or the Kruskal-Wallis test.

Results: Three distinct classes were obtained from the models constructed with LPA (Akaike information criteria: 1674.91, Bayesian information criteria: 1726.87, Lo-Mendell-Rubin adjusted likelihood ratio test: $p < 0.001$). The number of countries in the classes was 11, 14, and 4, respectively. The number of MRI units was the most prominent variable in separating the classes ($p = 0.001$). Türkiye was in the same class as Canada, Chile, the Czech Republic, Estonia, Hungary, Israel, Luxembourg, Mexico, Poland, and Slovenia. The numbers for all indicators in Türkiye were below the average of its class, except for the numbers of MRI units and medical graduates.

Conclusions: This study found the number of MRI units to be the most prominent indicator in categorizing OECD countries into three different classes, whereas the number of hospital beds and nurses did not differ across the defined classes.

Introduction

Healthcare systems play an active role in the development and welfare of a country. Social and economic conditions and health policies influence access to healthcare systems. Health indicators are important tools for monitoring and improving health systems. Governments use health indicators to guide healthcare system policies, set targets for improving population health, monitor public health programs, and make efficient plans. Researchers also use health indicators as supporting evidence to describe the state of a population's health (1). Indicators such as mortality rate, disease prevalence, and life expectancy can provide quantitative data to evaluate progress in healthcare systems. Health indicators can also highlight health inequalities

in a community. Differences between genders, races, ethnic groups, socioeconomic classes, and other groups in health indicators can be used to guide policies and interventions to bring about health equity in the future. Not only governments or researchers but also many institutions, including international organizations such as the United Nations, the World Health Organization, and the Organization for Economic Cooperation and Development (OECD), utilize health indicators (2).

The OECD is an international organization that began its activities in 1961 and includes 38 members. It works with governments to establish evidence-based international standards and find solutions to social, economic, and environmental challenges. Its objectives include improving



economic performance, creating jobs, promoting stronger education, exchanging experiences, advising on public policies, and setting international standards (3). In line with these objectives, it collects and shares data from member countries regularly. Health indicators have an important place in the data categories that include main headings such as agriculture, development, economy, education, and finance. OECD publishes various indicators of countries under the headings of health care utilization, health equipment, health resources, health risks, and health status.

Evaluating the efficiency of healthcare systems is crucial for resource allocation, cost control, quality improvement, equitable access, evidence-based decision-making, and international comparisons. In addition, it may support the optimal use of healthcare resources, improve quality of care, promote equitable access, make informed decisions, and learn from global best practices. It would be beneficial to classify OECD countries into homogenous subgroups using various health indicators for focused policy interventions, benchmarking and comparison, resource allocation, and progress monitoring. Previous studies have tested data envelopment and panel tobit methods to assess the efficiency of healthcare systems by grouping various health indicators as input and output variables (4-6). Cluster analysis (7-12), Application of Additive Ratio Assessment (ARAS), and Simple Additive Weightings (SAW) (13) methods have also been used to identify homogenous classes of member countries.

This study aimed to (i) allocate OECD countries into similar sub-classes using latent profile analysis (LPA) according to the number of health equipment and healthcare workers, (ii) determine the health indicators that are effective in the formation of the classes, and (iii) evaluate the status of Türkiye in its class.

Methods

Data Set

Data were collected for 38 OECD members, creating and collecting nine variables for each, covering the most recent values between 2018 and 2022. Open-access databases from the OECD Statistics and the World Bank were used (3,14). The dataset includes the number of hospital beds, computed tomography (CT) scanners, magnetic resonance imaging (MRI) units, mammography machines, and radiotherapy machines as health equipment indicators, and the numbers of doctors, nurses, medical graduates, and nursing graduates as healthcare workers indicators. The number of health equipment is presented per million inhabitants. The numbers of doctors and nurses are presented per thousand residents, whereas the numbers of graduates are presented per hundred thousand residents. Nine OECD countries, namely Costa Rica, Colombia, France, Germany, Japan, the Netherlands, Portugal, Switzerland, and the United Kingdom, were excluded from the analysis because

of missing data for one or more indicators. Previous authors also reported issues in data availability for some indicators and members of OECD countries (10,13,15).

Statistical Analysis

LPA was used to identify classes of OECD countries with similar health equipment and healthcare workers indicators. It is a person-centered statistical method that uses individual differences and similarities above relationships between variables and aims to divide individuals from a heterogeneous population into smaller, more homogeneous subgroups or classes based on continuous indicator variables (16).

LPA classifies observations using building models and is most interested in two parameters: (i) class or profile membership probabilities describing the class prevalence in the sample and (ii) means and variances of the indicator variables within each class. A widely used model equation for LPA is:

$$\sigma_i^2 = \sum_{k=1}^K \pi_k (\mu_{ik} - \mu_i)^2 + \sum_{k=1}^K \pi_k \sigma_{ik}^2,$$

In this formula, μ_{ik} and σ_{ik} indicate profile-specific (k) means and variances for indicator variable i , and π_k denotes the proportion of the individuals belonging to profile k .

LPA operates under the following assumptions: (i) samples drawn from a heterogeneous population produce data that are a mixture of K profile-specific distributions, (ii) each indicator variable follows a normal distribution, (iii) the profile-specific mean vectors represents the observed variable means for each profile (k), and (iv) indicator variables are uncorrelated within the identified latent profiles.

Determining the number of profiles (classes) (k) is a crucial step in LPA. Expectations of researchers, clinical relevance, theories generated from earlier studies, or statistical methods may all be considered when determining the number of profiles (17). Choosing the best model is a gradual process. The first step is to build a one-class LPA model that calculates the observed item proportions in the sample. This one-class model is a comparative baseline point for models with multiple classes. In the next step, the number of classes is increased by one, and the result is evaluated to determine whether the new solution outperforms the prior one conceptually and significantly. Building new models stops when convergence issues arise, or insufficient data exists to estimate all the model parameters (18).

Evaluation indices are used to select the best model among the models created. We employed the Akaike information criteria (AIC) (19), Bayesian information criteria (BIC) (20), and Lo-Mendell-Rubin adjusted likelihood ratio test (LMRT) (21) in this study. AIC and BIC are goodness-of-fit indices used to compare models. Lower values of AIC and BIC indicate a better fit of the model on the dataset (19,20). The LMRT compares the fit of a

target model with that of a model with one less class. A p-value <0.05 indicates that the target model fits significantly better (16).

Once the best model is determined, (i) researchers can investigate associations between indicator variables and the classes and, (ii) assign an individual to the most likely class based on posterior probabilities. Classes can be labeled according to the mean of each indicator variable in the profile (16,17). Those with less than 5% in the sample are misleading and tend to overfit the data (22).

All calculations were performed with Stata 18 (23), R studio (24), and the tidyLPA (Easily perform LPA Using Open-Source or Commercial Software) package (v1.1.0) (25). Parameter estimations of the LPA models were made with the Expected Maximization algorithm. Descriptive statistics of health indicators are displayed as mean \pm standard deviation and median (Q_1 - Q_3). The normality of the health indicators was assessed using the Shapiro-Wilk test. Comparison of classes based on health indicators was performed using one-way analysis of variance (ANOVA) and Kruskal-Wallis test for normally and non-normally distributed data, respectively. Pairwise comparisons were performed with the Bonferroni test for ANOVA, while Dunn's test was used for significant Kruskal-Wallis test results. A p value <0.05 was considered significant.

Results

We built a series of LPAs from one class to five. The AIC, BIC, and LMRT results are shown in Table 1. As more classes were introduced, the values of AIC and BIC decreased until the three latent-class solutions. Evaluation indices indicated the choice of the three-latent-class model by rejecting the one-

latent-class and two-latent-class models. The three-latent-class model was selected as the final model because it provided the best fit to the dataset.

The posterior probabilities of the members were calculated based on the three-class model. Class 1 countries had 11 (38%) members, while Class 2 and Class 3 countries had 14 (48%) and 4 (14%) members, respectively. The members of the classes are listed in Table 2. Türkiye was in Class 1 countries with countries from different regions. Class 2 countries mostly included members from European Union countries, whereas Class 3 countries included members from geographically diverse countries.

The health indicator values of the estimated latent classes were evaluated. Table 3 shows the comparisons of the estimated classes. The number of hospital beds ($p=0.990$) and nurses ($p=0.076$) did not significantly differ between the classes. Concerning the health equipment indicators, Class 1 countries had the lowest numbers of CT scanners ($p=0.001$), mammography ($p<0.001$), and radiotherapy units ($p=0.003$). The numbers of MRI units were significantly different among all classes ($p=0.001$). The highest number of MRI units belonged to Class 3 countries, followed by Class 2 countries ($p<0.001$) and Class 1 countries ($p<0.001$) (Table 3). Pairwise comparisons indicated no significant differences between the Class 2 and Class 3 countries in the number of CT scanners, mammography, and radiotherapy ($p>0.05$). Concerning the relationship between the estimated classes and indicators of healthcare workers, Class 2 countries had significantly higher numbers of doctors ($p=0.004$) and medical graduates ($p=0.010$) than Class 1 countries. Class 3 countries had the highest number of nursing graduates among the classes, whereas Class 1 and Class 2 countries had similar values ($p=0.011$).

Table 1. Comparisons of LPA models with different number of classes

Number of classes	Log Likelihood statistics	AIC	BIC	LMRT (p value)
1	-846.9345	1729.869	1754.480	-
2	-819.6528	1695.306	1733.590	<0.001
3	-799.4537	1674.907	1726.865	<0.001
4	-791.1796	1678.359	1743.989	0.130
5	-784.9831	1685.966	1765.269	0.336

AIC: Akaike information criterion, BIC: Bayesian information criterion, LMRT: Lo-Mendell-Rubin test

Table 2. Distribution of OECD countries in estimated classes

Class 1 countries (n=11)		Class 2 countries (n=14)		Class 3 countries (n=4)
Canada	Mexico	Australia	Lithuania	Finland
Chile	Poland	Austria	Latvia	Greece
Czech Republic	Slovenia	Belgium	Norway	Korea
Estonia	Türkiye	Denmark	New Zealand	The United States of America
Hungary		Ireland	Slovak Republic	
Israel		Iceland	Spain	
Luxembourg		Italy	Sweden	

OECD: Organization for Economic Co-operation and Development

Finally, we derived the estimated marginal means of the classes from the three-class LPA model. Values of all health indicators were converted to a scale of 0-100 for a better trend comparison on the latent profile plot shown in Figure 1. Class 3 countries showed a better trend for health equipment indicators while losing superiority in healthcare workers' indicators. Class 1 countries presented the lowest figures for all indicators.

Discussion

This study classified the OECD countries based on health indicators using LPA. Although there are different groups of indicators under the health heading on the OECD website, we prioritized the indicators of health equipment and healthcare workers to examine the status of healthcare systems. Some studies based on the classification of OECD countries focused mainly on indicators such as the number of nurses, doctors, hospital beds (7,8,10,11), health expenditure (7-9,11), or health

Table 3. Comparison of estimated classes by health indicators

Health indicators	Class 1 countries	Class 2 countries	Class 3 countries	P
	Mean±SD Median (Q ₁ -Q ₃)	Mean±SD Median (Q ₁ -Q ₃)	Mean±SD Median (Q ₁ -Q ₃)	
Hospital beds	3.99±1.92 4.18 (2.55-6.19)	4±1.59 3.3 (2.84-5.51)	5.62±4.73 3.51 (2.82-8.42)	0.990
CT scanners	16.16±5.62 ^A 16.26 (9.82-20.31)	34.05±14.01 ^B 29.69 (22.96-43.66)	35.99±12.73 ^B 41.61 (28.8-43.19)	0.001
MRI unit	10.43±4.47 ^A 11.41 (5.34-13.28)	18.14±7.42 ^B 16.33 (14.31-19.1)	34.17±2.92 ^C 33.9 (32.23-36.12)	0.001
Mammography	12.52±2.57 11.28 (10.78-14.79) ^A	20.46±7.63 17.37 (16.34-21.1) ^B	58.87±18.82 66.94 (47.98-69.76) ^B	<0.001
Radiotherapy	4.94±2.34 4.7 (2.77-6.64) ^A	9.39±3.45 8.2 (7.32-11.75) ^B	8.82±2.37 8.64 (6.8-10.83) ^B	0.003
Doctors	3.12±0.57 ^A 3.3 (2.7-3.48)	4.18±0.64 ^B 4.19 (3.67-4.48)	3.7±1.79 ^{A,B} 3.06 (2.52-4.89)	0.004
Nurses	6.83±3.14 6.38 (4.3-10.06)	10.31±4 10.67 (6.6-12.26)	9.32±4.53 10.18 (5.87-12.78)	0.076
Medical graduate	10.49±4.64 ^A 11.37 (7.55-14.28)	16.73±4.71 ^B 15.69 (13.57-21.04)	10.37±2.97 ^{A,B} 10.37 (7.88-12.85)	0.010
Nursing graduate	32.12±14.85 ^A 30.19 (18.48-37.36)	42.69±26.83 ^A 37.01 (24.27-46.15)	118.44±64.05 ^B 97.87 (78.54-158.35)	0.011

^{A, B, C}: Different letters on the same line represent significant differences (p<0.05).
SD: Standard deviation, CT: Computed tomography, MRI: Magnetic resonance imaging

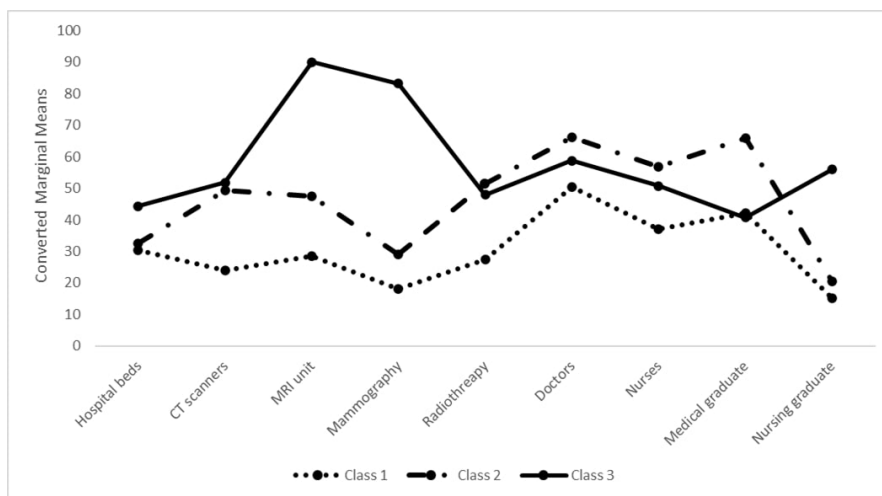


Figure 1. Latent profile plot of estimated classes based on the health indicators
CT: Computed tomography, MRI: Magnetic resonance imaging

status indicators such as infant mortality rates and life expectancy (8,12). However, few studies have been interested in indicators of health equipment (13). In this study, we included more health equipment indicators in the dataset. Unlike the previous studies, we focused on the indicators of both healthcare workers and health equipment to assess the current healthcare system status provided by OECD countries.

Another difference in the current work was utilizing the LPA during the classification process. To our knowledge, this is the first study classifying the OECD countries with LPA using health indicators. Unlike conventional classification methods (e.g., k-means clustering, hierarchical clustering), this method provides more objective and probability-based inferences (16,17). LPA estimates probabilities directly from the model to identify class membership and produces estimated marginal means to provide insight into the role of variables in creating classes (16-18). In addition, the appropriate number of classes is selected in a more realistic way using statistical methods such as LMRT (21).

Three distinct classes were derived from 29 countries, and there were significant differences between the estimated classes in health indicators, except for the number of hospital beds and nurses. The most prominent variable in separating the classes was the number of MRI units. These differences between estimated classes suggest that OECD countries are a heterogeneous population and that three distinct classes were created successfully. Using LPA in this study was an appropriate choice because it discriminates heterogeneous populations into more homogenous subclasses (16-18).

Türkiye was classified as a member of Class 1 countries in the study, while Canada, Chile, Czech Republic, Estonia, Hungary, Israel, Luxembourg, Mexico, Poland, and Slovenia were the others. This result is consistent with previous studies that classified OECD countries by health indicators. In these studies, Türkiye was placed in the same class as the Czech Republic, Chile, Israel, Mexico, Estonia, and Poland (8,10-12). The numbers of both health equipment and healthcare workers per capita in Class 1 countries were significantly lower than in other classes. Assessment of the current demographic, population growth rate and provision of health services may be an appropriate starting point to enhance healthcare systems in countries that require improvement. New policies to increase the quantity of healthcare workers and equipment may be subject to further review. An assessment of Türkiye's status in Class 1 countries revealed that the number of medical graduates and MRI units was higher than the average of its class. However, the remaining health indicators were below the class average. This situation illustrates that implementing new policies to increase the amount of health equipment and healthcare workers may enhance Türkiye's standing among OECD countries. Class 2

countries may be considered the most homogeneous class in the study because it included European Union countries, Australia, and New Zealand. Class 3 countries were the group with the least number of countries from the most diverse geographical regions. The common feature of Class 3 countries was that they included countries with high numbers of MRI units and nursing graduates. The highest number of MRI units among the OECD countries belonged to the United States of America, followed by South Korea and Greece. Finland had the highest number of nursing graduates, followed by Australia and South Korea.

The lack of data on health indicators for some countries in the OECD and World Bank databases could lead to bias in the study results.

Conclusion

This study categorized OECD countries into three classes. The number of MRI units was the most prominent indicator used in categorizing OECD countries into three classes. The number of hospital beds and nurses did not differ across the defined classes.

Ethics

Ethics Committee Approval: Ethical approval was not required.

Informed Consent: Informed consent was not required.

Authorship Contributions

Concept: H.Ö., D.Ö., Design: H.Ö., D.Ö., Data Collection or Processing: H.Ö., Analysis or Interpretation: H.Ö., D.Ö., Literature Search H.Ö., Writing: H.Ö., D.Ö.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Skolnik R. *Global Health 101*. 3rd ed. Burlington, USA: Jones & Bartlett Learning; 2016.
2. Pan American Health Organization. *Health Indicators - Conceptual and operational considerations*. Last Accessed Date: 01.08.2023. Available from: <https://www.paho.org/en/documents/health-indicators-conceptual-and-operational-considerations>.
3. Organization for Economic Cooperation and Development (OECD). Last Accessed Date: 01.08.2023. Available from: <https://www.oecd.org/about/>.
4. Kaya SP, Cafı R. Analysis of the efficiency determinants of health systems in OECD countries by DEA and panel tobit. *Soc Indic Res*. 2016;129:113-132.

5. Kocaman AM, Mutlu ME, Bayraktar D, Araz ÖM. Health care system efficiency analysis of OECD countries. *Engineer & the Machinery Magazine*. 2012;23:14-31.
6. Yüksel O. Comparison of healthcare system performances in OECD countries. *Int J of Health Serv Res and Policy*. 2021;6:251-261.
7. Alkaya A, Alkaş C. Cluster analysis classification of OECD countries according to health indicators. *Sosyal Güvence*. 2021;19:427-474.
8. Köse A. Evaluation of OECD countries according to clustering analysis method in terms of health indicators. *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi*. 2022;12:2010-2021.
9. Çetintürk İ, Gençtürk M. The classification of the health expenditure indicators of OECD countries through clustering analysis. *Süleyman Demirel University Visionary Journal*. 2020;11:228-244.
10. Dağtekin G, Kılınç A, Colak E, Ünsal A, Arslantas D. Classification of OECD countries in terms of medical resources and usage with Hierarchical Clustering Analysis. *Osmangazi Journal of Medicine*. 2022;44:487-492.
11. Değirmenci N, Yakıcı Ayan T. Evaluation of OECD countries according to fuzzy clustering analysis and topsis method in terms of health indicators. *Hacettepe University Journal of Economics and Administrative Sciences*. 2020;38:229-242.
12. Sonğur C. Cluster Analysis of Organization for Economic Cooperation and Development Countries According to Health Indicators. *Journal of Social Security*. 2016;6:197-224.
13. İzgüden D, Sezer Korucu K, Çalışkan Söylemez Ş, Demir M. The assessment of health indicators and health equipment of OECD countries with the entropy based aras and saw methods. *Süleyman Demirel University Visionary Journal*. 2022;13:731-755.
14. The World Bank. Health Nutrition and Population Statistics. Last Accessed Date: 01.08.2023. Available from: <https://databank.worldbank.org/source/health-nutrition-and-population-statistics>
15. Proksch D, Busch-Casler J, Haberstroh MM, Pinkwart A. National health innovation systems: clustering the OECD countries by innovative output in healthcare using a multi indicator approach. *Res Policy*. 2019;48:169-179.
16. Mathew A, Doorenbos AZ. Latent profile analysis—an emerging advanced statistical approach to subgroup identification. *Indian Journal of Continuing Nursing Education*. 2022;23:127-133.
17. Spurk D, Hirschi A, Wang M, Valero D, Kauffeld S. Latent profile analysis: a review and “how to” guide of its application within vocational behavior research. *J Vocat Behav*. 2020;120:103445.
18. Nylund-Gibson K, Choi AY. Ten frequently asked questions about latent class analysis. *Transl Issues Psychol Sci*. 2018;4:440-461.
19. Akaike H. Factor analysis and AIC. *Psychometrika*. 1987;52:317-332.
20. Schwarz G. Estimating the dimension of a model. *Ann Stat*. 1978;6:461-464.
21. Lo Y, Mendell NR, Rubin DB. Testing the number of components in a normal mixture. *Biometrika*. 2001;88:767-778.
22. Kircanski K, Zhang S, Stringaris A, et al. Empirically derived patterns of psychiatric symptoms in youth: a latent profile analysis. *J Affect Disord*. 2017;216:109-116.
23. StataCorp. Stata Statistical Software: Release 18. College Station, TX: StataCorp LLC. 2023.
24. R Team. RStudio: Integrated development environment for R.(Version 2022.7. 1.554). PBC, Boston, MA. Last Accessed Date: 01.06.2023. Available from: <http://www.rstudio.com>
25. Rosenberg JM, Beymer PN, Anderson DJ, Van Lissa CJ, Schmidt JA. tidyLPA: an R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *Journal of Open Source Software*. 2019;3:978.