

Clustering of OECD countries by chronic disease burden: identifying associated health system and socioeconomic factors via machine learning

Seda Sönmez

Department of Health Management, Faculty of Health Sciences, Kırıkkale University, Kırıkkale, Türkiye

Cite this article as: Sönmez S. Clustering of OECD countries by chronic disease burden: identifying associated health system and socioeconomic factors via machine learning. *J Med Palliat Care*. 2026;7(5):1185-1195. doi:10.47582/jompac.1961328

Received: 01.06.2026

Accepted: 22.10.2026

Published: 25.10.2026

ABSTRACT

Aims: The aim of this study is to cluster OECD (Organisation for Economic Co-operation and Development) countries based on their chronic disease burden and to rank the health system and socioeconomic variables that discriminate these clusters using machine learning.

Methods: Secondary data obtained from international databases were analyzed in two stages for 14 variables across 38 OECD countries. First, a K-means clustering analysis was conducted based on age-standardized DALY (disability adjusted life years) rates for five major chronic disease groups and two associated risk factors. The statistical significance of cluster differences was tested using the Kruskal-Wallis test and the Bonferroni-corrected Dunn test. Second, health system and socioeconomic variables discriminating between clusters were ranked using the standard and balanced Random Forest models.

Results: OECD countries were grouped into three clusters: cluster 1 (n=4; Türkiye, United States, Mexico, Costa Rica), which exhibits a high burden of diabetes and kidney diseases and chronic respiratory diseases along with significant income inequality; cluster 2 (n=8; Central and Eastern Europe, Greece), which has the highest out-of-pocket payment rates along with a high burden of cardiovascular disease and tobacco-related risks; and cluster 3 (n=26; including Western European countries, Japan, Korea), which combines a relatively low burden of chronic diseases with high levels of health expenditure. Six out of seven health system and socioeconomic variables showed significant differences across the clusters. The standard Random Forest model identified current health expenditure and per capita income as the most discriminating variables, whereas in the balanced model, which accounted for class imbalance, income inequality ranked first.

Conclusion: The profiles of the burden of chronic diseases in OECD countries are associated with their socioeconomic and health system characteristics, particularly income inequality.

Keywords: Chronic disease, cluster analysis, machine learning, delivery of health care, socioeconomic factors, OECD

INTRODUCTION

In Organisation for Economic Co-operation and Development (OECD) countries, chronic diseases account for a significant majority of the total disease burden and are responsible for nearly 90% of all causes of mortality in European Union (EU) member states.^{1,2} These diseases impose an indirect burden on the overall economy through disability, early retirement, and lost productivity, in addition to the direct medical costs they impose on health systems. For example, it is estimated that the four major chronic diseases (cardiovascular diseases, cancer, diabetes, and chronic respiratory diseases) account for at least 25% of the total health budget in the EU and result in an economic loss equivalent to approximately 2% of gross domestic product.¹

One of the most critical factors driving the rising burden of chronic disease is the aging of the population; the proportion of the population aged 65 and older in OECD countries is projected to rise to 27.5% by 2050, and population ageing is

associated with a rise in chronic diseases.² Lifestyle-related risk factors such as tobacco and alcohol consumption, unhealthy diet, and physical inactivity also lead to metabolic issues that trigger chronic diseases like obesity, high blood pressure, and high blood sugar.^{3,4} Additionally, environmental factors such as socioeconomic inequalities, low income levels, lack of education, rapid urbanization, and air pollution, as well as workplace stress and long working hours, are among the key factors contributing to the disease burden.⁵ Furthermore, the financial structure of the healthcare system plays a significant role in the burden of chronic diseases. While increases in total health expenditures and the scope of public insurance reduce the burden of chronic diseases, rising out-of-pocket expenditures increase this burden.^{6,7} The management and prevention of chronic diseases are of great importance not only because they reduce individual quality of life and life expectancy but also because they increase healthcare

Corresponding Author: Seda Sönmez, sedakaya@kku.edu.tr



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

expenditures, particularly in cases of multimorbidity, and slow economic growth.^{7,8}

Although OECD countries appear to have similar levels of economic development, they exhibit a heterogeneous structure in terms of the design of their health systems, financing structures, prevalence of chronic diseases, and social conditions. Therefore, evaluating countries based on similar disease burden profiles rather than individual indicators increases the likelihood that interventions successful in one country will also be effective in other countries with similar profiles and helps identify inefficiencies in resource allocation.^{2,9} However, previous studies that have used country clustering have mostly focused on European countries and have examined the disease burden in conjunction with risk factors and disease management indicators.^{9,10} Based on this, the aim of this study is to cluster 38 OECD member countries solely based on their chronic disease burden profiles and to use machine learning to rank the health system and socioeconomic indicators that best distinguish these clusters. Thus, the study reveals not only which countries share similar disease burden profiles but also which health system and socioeconomic characteristics are associated with these profiles.

METHODS

Ethics

Ethics committee approval was not required for this study because it was based exclusively on publicly available, aggregated, and anonymized country-level secondary data and did not involve human participants or identifiable personal information. However, the entire research process was conducted in accordance with the principles of scientific research and publication ethics.

Research Design

This study is descriptive and analytical in nature and has a cross-sectional comparative design based on publicly available, country-level aggregated secondary data from 38 OECD countries. In the study, OECD countries were classified using cluster analysis based on chronic disease burden indicators; subsequently, the socioeconomic and health system indicators that differentiated these clusters were ranked by importance using the random forest method, a machine learning model.

Dataset and Variables

The study's population consists of all 38 OECD member countries and no country was excluded from the analysis. **Table 1** lists the variables used in the study, the sources of the data, the reference years, and the units of measurement.

Data on the burden of chronic disease were obtained from the Global Burden of Disease (GBD) 2023 study,¹¹ health system indicators from the OECD Health Statistics database,¹² and socioeconomic indicators from the World Bank's World Development Indicators database.¹³ The variables used in the analysis were categorized into three groups. The first group includes chronic disease burden indicators: DALY (disability-adjusted life years) rates for five major chronic disease groups, namely the four main noncommunicable disease groups (cardiovascular diseases, neoplasms, chronic respiratory diseases, diabetes and kidney diseases),¹ and musculoskeletal disorders (all GBD level 2 cause groups), which form the basis of the clustering analysis, as well as DALY rates attributable to two major modifiable risk factors, high BMI and smoking. Risk factor-attributable DALY rates were included in the analysis because they reflect the preventable component of the disease burden and represent areas where countries are open to policy intervention. The second and third groups include

Table 1. Research variables

Variables	Source	Year	Unit
A. Chronic disease burden (clustering variables)			
Cardiovascular diseases (CVD) DALY rate	GBD	2023	DALY/100,000
Diabetes and kidney diseases DALY rate	GBD	2023	DALY/100,000
Neoplasms (cancer) DALY rate	GBD	2023	DALY/100,000
Chronic respiratory diseases (CRD) DALY rate	GBD	2023	DALY/100,000
Musculoskeletal disorders (MSK) DALY rate	GBD	2023	DALY/100,000
High body-mass index (BMI) risk DALY rate	GBD	2023	DALY/100,000
Smoking risk DALY rate	GBD	2023	DALY/100,000
B. Health system (explanatory variables)			
Current health expenditure	OECD	2023	% of GDP
Out-of-pocket (OOP) expenditure rate	OECD	2023 ^a	% of current health expenditure
Preventive health expenditure	OECD	2023 ^b	% of current health expenditure
C. Socioeconomic (explanatory variables)			
GDP per capita (PPP)	World Bank	2023	Current international \$
Population aged 65 and above	World Bank	2023	% of total population
Urbanization rate	World Bank	2023	% of total population
Gini coefficient	World Bank	2019-2023 ^c	Index (0-100)

^a 2023 data were used for all countries except Norway (2022). ^b 2023 data were used for 33 countries and 2022 data for Israel and Norway. No data were available for 2019–2023 for three countries (COL, NZL, TUR; 7.9%), for which median imputation was applied. ^c For the Gini coefficient, the most recent available year was used (24 countries: 2023; 6 countries: 2022; 3 countries: 2021; 2 countries: 2020; 1 country: 2019). No data was available for two countries (HUN, NZL; 5.3 %), for which median imputation was applied. DALY: Disability-adjusted life years, GBD: Global Burden of Disease

variables related to socioeconomic factors and the health system to characterize and discriminate between the clusters. The year 2023 was used as the reference year for all variables, partly to minimize the exceptional impact of the COVID-19 pandemic on health expenditure. All GBD indicators and most explanatory variables refer to 2023; for countries without 2023 data, the most recent year between 2019 and 2023 was used ([Table 1](#)). Missing values remained in only two explanatory variables, preventive health expenditure (three countries, 7.9%) and the Gini coefficient (two countries, 5.3%), and were imputed with the median of the observed values across the 38 countries. Median imputation was preferred for three reasons. First, the proportion of missing data was low and confined to explanatory variables, leaving the clustering unaffected. Second, complete-case analysis would have excluded four countries (10.5%), including Türkiye, reducing cluster 1 to three countries. Third, the median is robust to skewed distributions and outliers. Multiple imputation was not used because it is unlikely to yield stable estimates in such a small sample.

Statistical Analysis

The aim of the study is not to estimate the effects on a single outcome variable, but rather to identify multidimensional disease burden profiles consisting of seven indicators and to rank the variables that best distinguish these profiles. To

accomplish this, a two-stage approach was adopted, and the data were analyzed accordingly. In the first stage, K-means clustering (non-hierarchical) analysis was performed using variables related to the burden of chronic disease. Z-score standardization was applied to ensure the comparability of variables with different scales and units. The analysis used the Hartigan-Wong algorithm based on Euclidean distance; it was run with 50 random starting points (nstart=50) and a maximum of 200 iterations (iter.max=200). A fixed random number of seeds (set.seed(42)) was specified to ensure reproducibility. The K-means method was chosen because it directly minimizes intra-cluster variance for a given number of clusters and, unlike agglomerative hierarchical methods, allows countries to be reassigned among clusters during iterations; the Ward method was used for sensitivity analysis ([Supplementary Table 1](#)). Additionally, the method's sensitivity to the choice of initial centers was reduced by using multiple random starting points. The optimal number of clusters was determined using two complementary criteria. The first criterion, the Elbow Method, examines the change in the within-cluster sum of squares (WCSS) for values of k ranging from 1 to 8. The second criterion, the average Silhouette width, measures the extent to which each country assigned to a cluster fits compared to its nearest neighbor's cluster for values of k ranging from 2 to 8. The number of clusters was determined as the value at which the Silhouette

Supplementary Table 1. Robustness of the three-cluster K-means results

A. Internal validity indices			
k	Calinski–Harabasz index	Gap (standard error)	Cluster sizes
2	16.42	0.327 (0.040)	11/27
3	17.53	0.324 (0.039)	4/8/26
4	16.82	0.329 (0.039)	1/9/9/19
5	15.93	0.316 (0.039)	1/3/7/8/19
6	15.65	0.333 (0.039)	1/3/5/7/7/15
7	15.09	0.329 (0.040)	1/3/4/4/5/7/14
8	15.25	0.346 (0.041)	1/2/2/4/4/5/6/14
B. Cluster-wise bootstrap stability (1,000 bootstrap resamples)			
Cluster	Countries (n)	Mean Jaccard similarity	Dissolved, n/1,000
Cluster 1	4	0.639	468
Cluster 2	8	0.862	43
Cluster 3	26	0.889	4
Two-cluster solution (k=2)	11/27	0.824/0.935	–
C. Agreement with alternative clustering algorithms			
Method	Adjusted Rand index	Differences from the K-means solution	
Partitioning around medoids (k=3)	0.918	Colombia grouped with cluster 1	
Ward's method (k=3)	0.688	Mexico formed a single-country cluster; the United States and Turkiye grouped with six cluster 2 countries; Czechia, Greece, and Costa Rica grouped with cluster 3	
D. Leave-one-country-out re-estimation (38 runs)			
Country excluded	Adjusted Rand index with original assignment		Cluster 1 retained
Any of the other 35 countries	1.000		Yes
Czechia	0.903		Yes
United States	0.970		No
Mexico	0.537		No
The gap statistic selected k=1 (gap 0.317; standard error 0.041). An average Jaccard similarity of ≥0.75 indicates a stable cluster, while a value of 0.60–0.75 indicates a pattern with uncertain membership; a cluster with a Jaccard similarity <0.50 is considered dissolved in that resample.			

The gap statistic selected k=1 (gap 0.317; standard error 0.041). An average Jaccard similarity of ≥ 0.75 indicates a stable cluster, while a value of 0.60–0.75 indicates a pattern with uncertain membership; a cluster with a Jaccard similarity < 0.50 is considered dissolved in that resample.

width reached its highest value and a distinct break in the WCSS curve was observed. The Calinski–Harabasz index and the gap statistic were also examined, and the stability of the three cluster solution was examined using bootstrap resampling, alternative clustering algorithms, and leave-one-country-out re-estimation ([Supplementary Table 1](#)).¹⁴ Differences between groups were identified using the Kruskal–Wallis test and the Bonferroni-corrected Dunn test. These comparisons were conducted solely for descriptive purposes, as disease burden indicators were used to define the clusters. Cluster centers were visualized using a heatmap.

In the second stage, the cluster assignments obtained in the first stage were defined as the dependent variable, and a Random Forest classification model was built using seven

health system and socioeconomic variables as explanatory variables. The number of potential variables evaluated at each node was determined as the square root of the number of explanatory variables, rounded to the nearest integer ($\sqrt{7} \approx 2.65$; mtry=3). Since the package's default setting rounds this value down (mtry=2), the results obtained with mtry=2 are also provided as a supplementary file ([Supplementary Table 2](#)). To reduce the risk of overfitting in the small sample, no hyperparameter tuning was performed, and other parameters were left at their default values (replacement bootstrap sampling, minimum terminal node size=1). Since preliminary analyses showed that the variable importance rankings obtained with 1,000 trees varied depending on the random number seed, the number of trees was set to 5,000

Supplementary Table 2. Out-of-bag confusion matrices and sensitivity analyses of the Random Forest models

A. Out-of-bag confusion matrices						
Actual cluster	Predicted cluster 1	Predicted cluster 2	Predicted cluster 3	Correct, % (95% CI)		
Standard model (mean of 50 runs)						
Cluster 1 (n=4)	1.1	0.0	2.9	28.0 (0.6–80.6)		
Cluster 2 (n=8)	0.0	5.0	3.0	62.5 (24.5–91.5)		
Cluster 3 (n=26)	2.5	1.9	21.6	83.1 (65.1–95.6)		
Overall				72.9 (56.9–86.6)		
Balanced model (identical in all 50 runs)						
Cluster 1 (n=4)	3	0	1	75.0 (19.4–99.4)		
Cluster 2 (n=8)	0	8	0	100.0 (63.1–100.0)		
Cluster 3 (n=26)	3	4	19	73.1 (52.2–88.4)		
Overall				78.9 (62.7–90.4)		
B. Variable importance ranks (Mean Decrease Gini) under alternative specifications						
Variable	Standard model			Balanced model		
	mtry=3	mtry=2	Complete case	mtry=3	mtry=2	Complete case
Current health expenditure	1	1	2	4	4	5
GDP per capita	2	2	1	5	5	4
Gini coefficient	3	4	4	1	1	1
Urban population	4	3	3	2	2	2
Population aged 65 and over	5	5	5	3	3	3
OOP expenditure	6	6	6	6	6	6
Preventive health expenditure	7	7	7	7	7	7
Accuracy, %	72.9	75.2	80.9	78.9	78.9	78.4
Cohen's Kappa	0.41	0.44	0.57	0.62	0.62	0.60
Most important for cluster 1	Gini	Gini	Gini	Gini	Gini	Gini
C. Stratified subsampling (200 subsamples, 80% of countries in each cluster)						
Variable	Standard model			Balanced model		
	Mean rank	Top 3, %	Top for cluster 1, %	Mean rank	Top 3, %	Top for cluster 1, %
Current health expenditure	1.44	100	18	3.61	28	0
GDP per capita	1.62	100	3	4.00	22	0
Gini coefficient	3.62	51	79	1.85	80	96
Urban population	3.88	38	0	2.88	74	0
Population aged 65 and over	4.91	8	0	2.75	96	4
OOP expenditure	5.53	3	0	6.11	0	0
Preventive health expenditure	7.00	0	0	6.82	0	0
Panel B shows the average ranks for 50 repetitions. Confidence intervals are exact binomial intervals (based on rounded average counts in the standard model); the majority class rate was 68.4%. Panel C shows the percentage of subsamples in which the variable ranks in the top three in the overall ranking or first in cluster 1. Complete case analysis: n=34 (Colombia, Hungary, New Zealand, and Türkiye excluded). In both the complete case and subsample analyses, the balanced model sampled two countries from each cluster. CI: Confidence interval. OOP: Out-of-pocket						

Panel B shows the average ranks for 50 repetitions. Confidence intervals are exact binomial intervals (based on rounded average counts in the standard model); the majority class rate was 68.4%. Panel C shows the percentage of subsamples in which the variable ranks in the top three in the overall ranking or first in cluster 1. Complete case analysis: n=34 (Colombia, Hungary, New Zealand, and Türkiye excluded). In both the complete case and subsample analyses, the balanced model sampled two countries from each cluster. CI: Confidence interval, OOP: Out-of-pocket

(ntree= 5,000). To account for the stochastic nature of the algorithm, the model was run 50 times with different random number seeds (1-50), and all results were reported as the mean and range (lowest–highest) of these 50 repetitions; this range reflects algorithmic variability only, not uncertainty due to small sample. Because there was a significant imbalance among the clusters (4, 8, and 26 countries), a balanced Random Forest model was also estimated, in which each tree was trained on a stratified sample consisting of three countries from each cluster, without replacement. The balanced model was also run with the same settings (ntree=5000, mtry=3, 50 random seeds). While the standard model reflects the actual distribution of countries across clusters, the balanced model is better suited for identifying the variables that distinguish small clusters because it assigns equal weight to each disease burden profile. Rather than selecting one of the models afterward, both were reported, and the conclusions were based primarily on findings that were consistent across both models. The uncertainty arising from the small sample size and the robustness of the findings were assessed using exact binomial confidence intervals, a complete case analysis excluding the four countries with imputed values, and 200 stratified subsamples (Supplementary Table 2). Variable importance indicates discriminatory ability rather than a causal effect. Since the sample size ($n=38$) was not sufficient to create separate training and test sets, model performance was evaluated using the out-of-bag (OOB) approach. Model performance was assessed using the OOB error rate, overall accuracy, Cohen's Kappa coefficient, and cluster-based correct classification rates (OOB confusion matrices are given in Supplementary Table 2). In addition, the countries misclassified in the OOB predictions and the frequencies of misclassification were examined over 50 iterations. A country was considered misclassified when the cluster to which it belonged, as predicted by the Random Forest model based on socioeconomic and health system indicators, differed from the cluster to which it actually belonged based on its disease burden profile. Variable importance was primarily determined using the Mean Decrease Gini (MDG) metric. This metric represents the average reduction in impurity achieved by a variable across all trees' node splits and is conceptually independent of the Gini coefficient, which serves as an indicator of income inequality in the dataset. To examine the cluster-specific importance of variables, the mean decrease accuracy (MDA) was also calculated. All analyses were performed in R (version 4.5.3; R Foundation for Statistical Computing, Vienna, Austria) using the cluster (2.1.8.2), factoextra (2.0.0), fpc (2.2.15), mclust (6.1.3), randomForest (4.7-1.2), caret (7.0-1), and dunn.test (1.3.7) packages.

RESULTS

The findings of the study are presented in this section using tables and figures. The Silhouette and Elbow methods were applied together to determine the optimal number of clusters. The Silhouette score reached its highest value (0.374) at $k=3$ (Table 2).

In the elbow plot, it was observed that the rate of decline in WCSS slowed down starting at $k=3$ (see Figure 1). The three-cluster solution also has the highest Calinski–Harabasz index;

Table 2. Average Silhouette scores

Number of clusters (k)	Average Silhouette score
2	0.368
3	0.374
4	0.292
5	0.259
6	0.238
7	0.216
8	0.231

the gap statistic, however, did not indicate a distinct cluster structure. Clusters 2 and 3 are stable across bootstrap samples and largely reproduced by alternative algorithms; the four-country cluster 1, however, is less stable and relies heavily on Mexico (Supplementary Table 1). In the two-cluster solution, cluster 2 and three of the four cluster 1 countries (all except Costa Rica) formed a single high burden group.

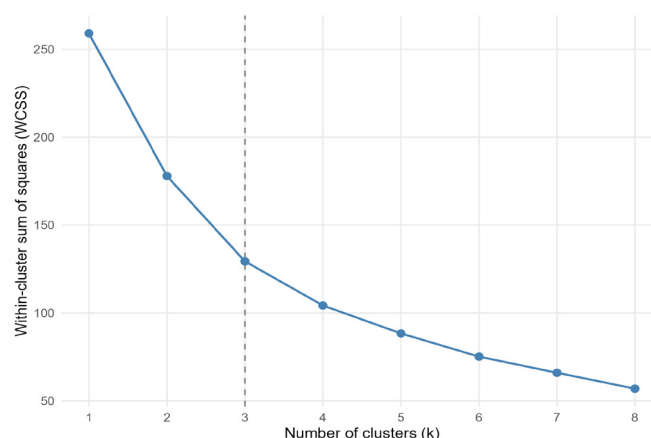


Figure 1. Optimal number of clusters (Elbow method) ($k=1$ to $k=8$, within-cluster sum of squares)

Accordingly, a K-means cluster analysis was conducted, and the 38 OECD countries were divided into three distinct groups based on their chronic disease burden profiles. The cluster assignments are presented in Table 3.

The median values for the burden of chronic disease, socioeconomic, and health system variables across the three clusters are presented in Table 4. The Kruskal-Wallis test conducted to determine differences between clusters identified statistically significant differences between clusters in 12 of the 14 variables ($p<0.05$). Musculoskeletal disorders ($p=0.412$) and preventive health expenditures ($p=0.526$) did not show significant cluster differences.

Cluster 1 reflects the metabolic and respiratory burden profile, with the highest median DALY rates for diabetes and kidney diseases (2,033.9 per 100,000), chronic respiratory diseases (1,127.2 per 100,000), and high BMI-attributable burden (2,335.7 per 100,000). For diabetes and kidney diseases, high BMI, and CVD, Cluster 1 did not differ significantly from cluster 2, but both clusters were significantly higher than cluster 3; for chronic respiratory diseases, cluster 1 was significantly higher than cluster 2 but did not differ significantly from cluster 3. The smoking-attributable burden

Table 3. Distribution of countries by clusters

Cluster	Profile	n	Countries
Cluster 1	Metabolic and respiratory burden profile	4	Costa Rica (CRI), Mexico (MEX), Türkiye (TUR), United States (USA)
Cluster 2	Cardiovascular and smoking-related burden profile	8	Czechia (CZE), Estonia (EST), Greece (GRC), Hungary (HUN), Latvia (LVA), Lithuania (LTU), Poland (POL), Slovakia (SVK)
Cluster 3	Relatively low chronic disease burden profile	26	Australia (AUS), Austria (AUT), Belgium (BEL), Canada (CAN), Chile (CHL), Colombia (COL), Denmark (DNK), Finland (FIN), France (FRA), Germany (DEU), Iceland (ISL), Ireland (IRL), Israel (ISR), Italy (ITA), Japan (JPN), Korea (KOR), Luxembourg (LUX), Netherlands (NLD), New Zealand (NZL), Norway (NOR), Portugal (PRT), Slovenia (SVN), Spain (ESP), Sweden (SWE), Switzerland (CHE), United Kingdom (GBR)

Table 4. Hypothesis test results by cluster

Variables	Cluster 1 (n=4)	Cluster 2 (n=8)	Cluster 3 (n=26)	χ^2	p
Chronic disease burden (DALYs per 100,000)					
CVD	3279.5 (3081.9-3554.9) ^a	4316.1 (3782.5-5609.5) ^a	1995.1 (1810.8-2265.1) ^b	23.2	<0.001
Diabetes and kidney diseases	2033.9 (1800.5-2653.6) ^a	978.1 (876.0-1138.6) ^a	719.4 (607.4-779.5) ^b	19.4	<0.001
Neoplasms	2908.6 (2726.7-2998.2) ^a	3701.6 (3230.6-3749.1) ^b	2902.6 (2668.0-3188.5) ^a	12.4	0.002
CRD	1127.2 (991.8-1211.6) ^a	560.3 (418.8-675.2) ^b	798.1 (650.6-918.6) ^a	11.6	0.003
MSK	2154.9 (2106.6-2332.6)	2058.9 (1989.5-2170.0)	2091.9 (1952.8-2195.3)	1.8	0.412
High BMI-attributable burden	2335.7 (2125.0-2696.7) ^a	1793.6 (1581.6-1994.1) ^a	1041.7 (929.8-1208.9) ^b	24.0	<0.001
Smoking-attributable burden	1172.9 (670.7-1737.0) ^{ab}	1932.3 (1881.6-2034.4) ^a	1132.1 (1030.1-1348.1) ^b	15.9	<0.001
Health system and socioeconomic indicators					
Current health expenditure (% of GDP)	6.28 (5.32-9.37) ^{ab}	7.32 (7.22-7.67) ^a	10.06 (8.71-10.79) ^b	9.9	0.007
OOP expenditure (% of CHE)	21.59 (17.09-27.87) ^{ab}	22.60 (19.08-31.84) ^a	14.57 (12.44-20.76) ^b	6.6	0.037
Preventive health expenditure (% of CHE)	2.86 (2.14-3.19)	2.78 (2.66-3.32)	3.12 (2.32-4.44)	1.3	0.526
GDP per capita, PPP (thousand international \$)	35.85 (27.99-52.67) ^{ab}	47.53 (45.16-49.56) ^a	63.82 (55.78-76.81) ^b	11.3	0.003
Population aged 65 and over (%)	10.86 (9.50-13.15) ^a	20.68 (19.78-21.02) ^b	19.88 (16.94-21.16) ^b	8.6	0.014
Urban population (%)	79.77 (79.33-82.31) ^{ab}	69.43 (66.24-71.19) ^a	83.12 (78.40-88.61) ^b	10.6	0.005
Gini coefficient (0-100)	43.60 (43.08-44.45) ^a	31.67 (27.80-33.55) ^b	32.35 (29.07-33.77) ^b	9.2	0.010

Values are medians (interquartile range, 25th–75th percentile). χ^2 and p values are from the Kruskal–Wallis test (df=2). Superscript letters indicate Dunn's post hoc comparisons with Bonferroni adjustment: clusters sharing a letter do not differ significantly (adjusted $p \geq 0.05$). Bonferroni-adjusted p-values for all pairwise comparisons are given in [Supplementary Table 3](#). Letters are not shown when the Kruskal–Wallis test was not significant. Missing values (preventive health expenditure: COL, NZL, TUR; Gini coefficient: HUN, NZL) were imputed with the median. DALY: Disability-adjusted life years, CVD: Cardiovascular diseases, CRD: Chronic respiratory diseases, MSK: Musculoskeletal disorders, BMI: Body-mass index, OOP: Out-of-pocket, CHE: Current health expenditure

Supplementary Table 3. Bonferroni-adjusted p-values for pairwise comparisons between clusters (Dunn's test)

Variable	Kruskal–Wallis p	Cluster 1 vs 2	Cluster 1 vs 3	Cluster 2 vs 3
Disease burden (DALYs per 100,000)				
CVD	<0.001	1.000	0.027	<0.001
Diabetes and kidney diseases	<0.001	0.596	<0.001	0.009
Neoplasms	0.002	0.034	1.000	0.002
CRD	0.003	0.003	0.217	0.034
MSK	0.412	0.742	0.573	1.000
High BMI-attributable burden	<0.001	1.000	<0.001	<0.001
Smoking-attributable burden	<0.001	0.053	1.000	<0.001
Health system and socioeconomic indicators				
Current health expenditure	0.007	1.000	0.172	0.015
OOP expenditure	0.037	1.000	0.659	0.044
Preventive health expenditure	0.526	1.000	1.000	1.000
GDP per capita	0.003	1.000	0.127	0.008
Population aged 65 and over	0.014	0.012	0.033	0.959
Urban population	0.005	0.136	1.000	0.004
Gini coefficient	0.010	0.014	0.012	1.000

Bonferroni-corrected p-values for the two-tailed, three pairwise comparisons obtained from the Dunn test; values less than 0.05 are shown in bold. DALY: Disability-adjusted life years, CVD: Cardiovascular diseases, CRD: Chronic respiratory diseases, MSK: Musculoskeletal disorders, BMI: Body-mass index, OOP: Out-of-pocket

in cluster 1 did not differ significantly from either of the other clusters. From a socioeconomic perspective, this cluster had the highest income inequality (median Gini coefficient 43.60) and the lowest proportion of the population aged 65 and over (10.86%), both significantly different from clusters 2 and 3.

Cluster 2, consisting mainly of Central and Eastern European countries, reflects the cardiovascular and smoking-related burden profile. This cluster had the highest median DALY rates for CVD (4,316.1 per 100,000) and smoking-attributable burden (1,932.3 per 100,000), both significantly higher than in cluster 3, as well as the highest neoplasm burden (3,701.6 per 100,000), which was significantly higher than in both other clusters. In contrast, its burden of chronic respiratory diseases was significantly lower than in both other clusters. Cluster 2 had the highest median out-of-pocket expenditure (22.60% of current health expenditure), and its health expenditure (7.32% of GDP), GDP per capita, and urbanization rate were significantly lower than those of cluster 3.

Cluster 3 exhibits a relatively low chronic disease burden profile and includes most Western, Northern, and Southern European countries, as well as Japan, Korea, Canada, Australia, New Zealand, Israel, Chile, and Colombia. DALY rates for CVD, diabetes and kidney diseases, and high BMI-attributable burden were significantly lower than in clusters 1 and 2, whereas the neoplasm burden was similar to that of cluster 1 and significantly lower than that of cluster 2. This cluster had the highest median health expenditure (10.06% of GDP) and GDP per capita (63.82 thousand international \$) and the lowest out-of-pocket expenditure (14.57%); these differences were statistically significant compared with cluster 2 but not with cluster 1.

K-means cluster centers were visualized using a Z-score heatmap (Figure 2). In Cluster 1, diabetes and kidney diseases ($Z=2.30$) and high BMI-attributable burden ($Z=2.04$) are markedly elevated, indicating a predominant metabolic

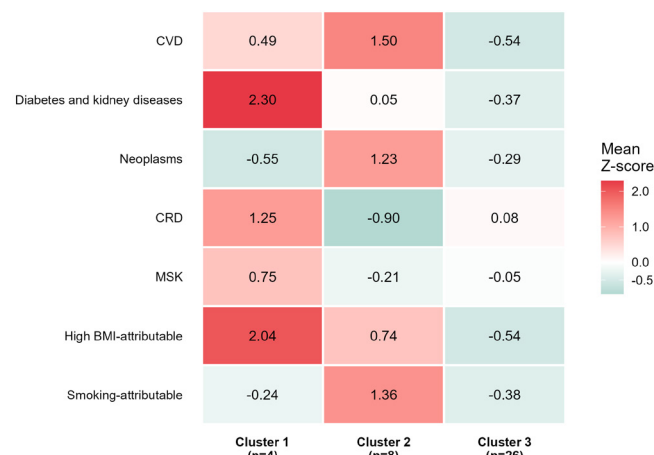


Figure 2. Heatmap for clusters (Heatmap of standardized Z-scores for seven DALY variables across three clusters)

DALY: Disability-adjusted life years, CVD: Cardiovascular diseases, CRD: Chronic respiratory diseases, MSK: Musculoskeletal disorders, BMI: Body-mass index

burden; together with the above-average CRD burden ($Z=1.25$), this supports the metabolic and respiratory burden profile. In Cluster 2, CVD ($Z=1.50$), smoking-attributable burden ($Z=1.36$), and neoplasms ($Z=1.23$) stand out with high values, whereas CRD is below average ($Z=-0.90$), which is consistent with the cardiovascular and smoking-related burden profile together with a high neoplasm burden. Cluster 3, on the other hand, reflects a generally low chronic disease burden with negative or near-zero Z-scores across all dimensions. Although cluster 1 showed an above-average Z-score for MSK ($Z=0.75$), this difference was not statistically significant in the Kruskal-Wallis test ($p=0.412$). These labels describe the dominant characteristics of each cluster center but do not imply that each indicator differs significantly across clusters (Table 4; Supplementary Table 3).

Table 5 presents the performance of the standard and balanced Random Forest models along with the ranking of variable importance. In the standard model, current health

Table 5. Variable importance and model performance for Random Forest

Variable	Standard Random Forest			Balanced Random Forest		
	Rank	MDG	Consistency (%)	Rank	MDG	Consistency (%)
A. Variable importance (MDG)						
Current health expenditure	1	4.48	100	4	0.87	98
GDP per capita	2	4.21	100	5	0.82	98
Gini coefficient	3	2.55	98	1	1.42	100
Urban population	4	2.43	98	2	1.16	100
Population aged 65 and over	5	2.07	100	3	1.04	100
OOP expenditure	6	1.32	100	6	0.41	100
Preventive health expenditure	7	0.56	100	7	0.27	100
B. Model performance (OOB)						
	Mean (min-max)			Mean (min-max)		
Overall accuracy, %	72.9 (71.1-76.3)			78.9 (78.9-78.9)		
Cohen's Kappa	0.41 (0.38-0.49)			0.62 (0.62-0.62)		
Correct classification, cluster 1, %	28.0 (25.0-50.0)			75.0 (75.0-75.0)		
Correct classification, cluster 2, %	62.5 (62.5-62.5)			100.0 (100.0-100.0)		
Correct classification, cluster 3, %	83.1 (80.8-88.5)			73.1 (73.1-73.1)		

Consistency is the percentage of runs in which the variable occupied the rank shown. MDG values are means across the 50 runs. MDG values are not directly comparable between the two models because of different sample sizes per tree; ranks are comparable. Predictions of the balanced model were identical in all 50 runs. MDG: Mean Decrease Gini, OOP: Out-of-pocket, OOB: Out-of-bag, Min: Minimum, Max: Maximum

expenditures and GDP per capita emerged as the most important variables, followed by the Gini coefficient and the urbanization rate. In the balanced model, the Gini coefficient, the urbanization rate, and the population aged 65 and older ranked in the top three, followed by health expenditures and GDP per capita. In both models, the Gini coefficient was the most important variable for cluster 1 in all trials, while out-of-pocket expenditures and preventive health expenditures ranked last. The standard model's average overall accuracy was calculated as 72.9% (71.1-76.3) and the Kappa coefficient as 0.41 (0.38-0.49), indicating moderate agreement; the overall accuracy only slightly exceeded the majority-class rate of 68.4%. The correct classification rates were 28.0%, 62.5%, and 83.1% for clusters 1, 2, and 3, respectively. The balanced model, which produced the same predictions across all runs, demonstrated substantial agreement, achieving 78.9% overall accuracy and a Kappa coefficient of 0.62; the correct classification rates were 75.0%, 100%, and 73.1% for clusters 1, 2 and 3, respectively. The United States (classified as cluster 3 instead of cluster 1), Colombia and Chile (classified as cluster 1 instead of cluster 3), and Portugal and Italy (classified as cluster 2 instead of cluster 3) were consistently misclassified in both models. These findings remained largely unchanged in the complete case analysis and in most stratified subsamples; the Gini coefficient continued to be the most important variable for cluster 1. However, the confidence intervals for the accuracy estimates are wide and include the majority class rate ([Supplementary Table 2](#)).

DISCUSSION

In this study, OECD countries were clustered based on their chronic disease burden profiles and were found to fall into three groups. This finding indicates that the burden of chronic disease is not distributed uniformly across countries and is not determined by economic level alone. For example, the United States shared a burden profile with middle-income countries such as Mexico, Costa Rica, and Türkiye, whereas Colombia and Chile, which belong to the low-burden cluster 3, had socioeconomic characteristics similar to those of cluster 1. Studies have also shown that OECD countries, particularly in Europe, are divided into different groups based on health status indicators (such as DALYs and life expectancy at birth), and that there are distinct differences between these groups in terms of behavioral patterns and geographic characteristics.⁹ Similarly, a recent study examining the EU-27+2 countries identified seven clusters. According to this study, the OECD member countries in the three clusters with the highest burden of cardiovascular disease are the same eight countries that form cluster 2 in this study, despite the use of different indicators and methods.¹⁰ In another study, among European countries in 2019, CVD accounted for about one-third of the chronic disease DALY rate in Latvia (32.1%), compared with 10.5–12.5% in France, Spain, and Iceland.⁴ This variability is also evident in conditions such as stroke, diabetes, and chronic respiratory diseases; the disease burden from stroke can be up to ten times higher in some regions compared to others.⁵ Compared to these analyses, this study groups all 38 OECD member countries based on their age-standardized

disease burden and associates the resulting profiles with health system and socioeconomic characteristics.^{4,9,15-18}

The fact that Türkiye, the United States, Mexico, and Costa Rica (cluster 1) exhibit a high burden of diabetes and kidney diseases and high BMI-attributable disease may reflect the “nutritional transition” triggered by globalization and rapid urbanization; and that this process has accelerated the shift toward out-of-home consumption, high-calorie fast-food-style diets, and sedentary lifestyles.¹⁵ In Türkiye, this transition is reflected in one of the highest rates of adult obesity in the WHO European Region (32.1%).¹⁹ While these countries are at different stages of the epidemiological transition, they share high income inequality. Cluster 1 has the highest Gini coefficient, and the Gini coefficient has been the most important variable for this cluster in all iterations of both Random Forest models. Socioeconomic analyses demonstrate a significant and strong association between risk factors for chronic diseases and income inequality, showing that the risk of obesity and diabetes spreads more rapidly in societies with high income inequality.^{16,20} Considering the imbalance in cluster sizes, the balanced Random Forest model correctly classified Türkiye, Mexico, and Costa Rica as cluster 1 based solely on socioeconomic and health system indicators. In contrast, although the United States belongs to cluster 1 in terms of disease burden, it was classified as cluster 3 by both models in all iterations. In other words, the United States' socioeconomic and health system indicators resemble those of the high-income countries in cluster 3. Consequently, the United States combines the economic and health system characteristics of high-income countries with a metabolic disease burden profile shared with middle-income countries; relatively high income inequality may be relevant to this pattern.

While the high burden of CVD and tobacco-related diseases in Eastern European countries in cluster 2 is well documented in the literature, this study shows that this is accompanied by lower health expenditures, GDP per capita, and urbanization rates, as well as a higher share of out-of-pocket payments, compared to cluster 3. In Random Forest models, the variables that best distinguished these two groups were health expenditure and GDP per capita in the standard model, and urbanization in the balanced model; the contribution of out-of-pocket payments remained limited. The concurrently high burden of neoplasms in cluster 2 is also consistent with this cluster's high burden attributable to smoking. Indeed, tobacco use is a common risk factor for cardiovascular diseases and many types of cancer.²¹ This finding is consistent with evidence that, in addition to traditional risk factors such as smoking, high blood pressure, unhealthy diets, and heavy alcohol consumption, difficulties in accessing health care due to inequities in access and insufficient investment in preventive health care contribute to the burden of preventable chronic diseases.²²⁻²⁴ The relatively low burden of chronic diseases in cluster 3, comprising mostly high-income countries, may reflect higher health expenditure and economic resources, as well as factors not measured in this study, such as healthy lifestyle habits and health system capacity.²⁵ Consistent with their higher per capita GDP, these countries allocate more

resources to their health systems, which may facilitate the prevention and management of chronic diseases. Additionally, healthy dietary patterns characterized by a variety of plant-based foods and the restriction of processed foods serve as a key protective factor in preventing chronic conditions such as cardiovascular diseases and diabetes.²⁶ This may be particularly true for Italy and Portugal, which, although predicted to be in cluster 2 by both models, actually belong to cluster 3 meaning their socioeconomic characteristics are closer to those of cluster 2.

The most important variables that differentiate countries based on their chronic disease burden profiles were ranked using two random forest models. In the standard model, health expenditure and GDP per capita were found to be the most important variables. This finding is consistent with studies reporting a relationship between health expenditure and health outcomes.^{27,28} However, these variables which are two closely related indicators of economic capacity primarily distinguished cluster 3 from cluster 2. In the balanced model, where equal weight was assigned to each disease burden profile, income inequality, urbanization, and population aging emerged as the most important variables. The balanced model classified small clusters considerably better and produced consistent predictions across repetitions. However, this led to a decrease in accuracy for cluster 3. Population aging primarily distinguished the younger populations of cluster 1. Since DALY rates were age-standardized, this relationship is not a direct result of differences in age structure; instead, it may reflect differences in the stage of demographic and epidemiological transition. Urbanization, on the other hand, primarily distinguished cluster 2, which had the lowest urbanization rate. The Gini coefficient, an indicator of income inequality, emerged as the most important variable for cluster 1 in all model specifications and in most subsamples; however, its ranking in the overall hierarchy varied depending on the model. This finding is consistent with the importance of social determinants of health, but it reflects an association rather than a causal effect. Research has shown that DALY rates increase in countries with high income inequality.^{9,16} This situation creates a “social gradient” in which individuals of lower socioeconomic status are more exposed to risk factors such as unhealthy diets, tobacco use, and physical inactivity, and face financial barriers to accessing preventive health services.^{4,29} Particularly in countries with pronounced income inequality, such as Türkiye, the United States, and Mexico, this pattern has been associated with high prevalence rates of obesity and diabetes.²⁰

Preventive health expenditure was the least important variable in both models and showed no significant differences across clusters. This finding should not be interpreted as preventive health expenditure is at a uniform level across OECD countries. Indeed, the share of these expenditures within current health expenditures ranges from 0.3% in Israel to 6.7% in Canada. Since this indicator is expressed as a share of current health expenditures, it reflects the relative priority given to preventive care rather than the absolute level of preventive spending. Therefore, countries with high total health expenditures may spend more per capita on preventive

services, even if the share of such spending is similar or lower. This finding primarily reflects significant variability within the clusters and certain measurement issues. First, the scope of preventive services in national health accounts varies across countries. Preventive services provided during primary care visits are often recorded as curative services, which can lead to an underestimation of preventive health expenditure.² The consistently low values reported by Israel for the 2019–2022 period (0.25%–0.32%) may reflect this reporting practice rather than actual differences in preventive activities. Second, expenditures on COVID-19 testing and vaccination were recorded as preventive services and may have inflated the figures for the pandemic years, including the 2022 values reported for Israel and Norway. Third, current preventive health expenditure may not reflect the disease burden that has accumulated over decades. Nevertheless, preventive health expenditure remains low in most countries. While the average share of preventive health expenditure in GDP (rather than in current health expenditure, as in the present study) across 36 OECD countries was 0.237% in 2017, it has been estimated that the share needed to minimize the prevalence of ill health and maximize economic growth and well-being is 1.175% of GDP.³⁰

Limitations

While this study fills an important gap in the literature, it also has certain limitations. First, due to its cross-sectional data structure, the causal nature of the relationships between variables could not be assessed. Given that the disease burden reflects cumulative exposures over decades, it is recommended that future studies examine temporal dynamics using panel data approaches. Second, the sample was limited to 38 countries, and there is a significant imbalance in cluster sizes and the cluster structure was only weakly separated, with the four-country cluster 1 being less stable than other clusters. Therefore, the classification performance of the standard Random Forest model remained moderate, the confidence intervals for accuracy were wide in both models, and the ranking of variable importance varied depending on whether cluster sizes were balanced or not. Furthermore, closely related variables such as health expenditure and GDP per capita may share importance in Random Forest models. Third, the reference year differs across countries for some explanatory variables. Although these differences are limited to a small number of countries and largely concern slowly changing indicators such as the Gini coefficient, the 2022 values for preventive health expenditure may partially reflect pandemic-related spending. Fourth, the median imputation applied to preventive health expenditure (7.9% missing) and the Gini coefficient (5.3% missing) may have reduced the variance of these variables and weakened differences across clusters although a complete-case analysis yielded consistent results. Finally, the GBD estimates for 2023 are model-based. Although each estimate is reported with a 95% uncertainty interval, only the central estimates were used in this study.

CONCLUSION

This study reveals that OECD countries can be divided into three clusters with different chronic disease burden profiles,

and that this cluster structure is associated with socioeconomic and health system factors, particularly income inequality, whereas the importance of health expenditure and per capita income depended on the modelling approach. The findings suggest that countries should adopt context-specific strategies aligned with their cluster profiles when determining policies related to chronic disease management. Türkiye's similarity to countries with a high burden of metabolic and respiratory diseases highlights the importance of prioritizing preventive interventions targeting obesity, diabetes, and tobacco control in health policies together with policies addressing income inequality. Future studies may be recommended to examine causal relationships and the impact of policy interventions on cluster profiles using panel data approaches.

ETHICAL DECLARATIONS

Ethics Committee Approval

Ethics committee approval was not required for this study because it was based exclusively on publicly available, aggregated, and anonymized country-level secondary data and did not involve human participants or identifiable personal information.

Informed Consent

Since this study is based solely on publicly available, aggregated, and anonymized country-level secondary data and does not include human participants or personally identifiable information, informed consent was not required.

Peer Review Process

This manuscript was subject to external peer review.

Conflict of Interest

The author declares no conflicts of interest related to this study.

Financial Disclosure

The author received no financial support for the conduct or publication of this research.

Author Contributions

The author is solely responsible for the entirety of conception, execution, analysis, and writing of the manuscript.

REFERENCES

- Vandenbergh D, Albrecht J. The financial burden of non-communicable diseases in the European Union: a systematic review. *Eur J Public Health*. 2020;30(4):833-839. doi:10.1093/eurpub/ckz073
- Gmeinder M, Morgan D, Mueller M. How Much Do OECD Countries Spend on Prevention? OECD Health Working Papers No 101. OECD Publishing; 2017. doi:10.1787/f19e803c-en
- Melaku YA, Renzaho A, Gill TK, et al. Burden and trend of diet-related non-communicable diseases in Australia and comparison with 34 OECD countries, 1990–2015: findings from the Global Burden of Disease Study 2015. *Eur J Nutr*. 2019;58(3):1299-1313. doi:10.1007/s00394-018-1656-7
- Andrade CAS, Mahrouseh N, Gabrani J, et al. Inequalities in the burden of non-communicable diseases across European countries: a systematic analysis of the Global Burden of Disease 2019 study. *Int J Equity Health*. 2023;22(1):140. doi:10.1186/s12939-023-01958-8
- Pomerleau J, Knai C, Nolte E. The burden of chronic disease in Europe. In: Nolte E, McKee M, eds. *Caring for People with Chronic Conditions: A Health System Perspective*. Open University Press; 2008:15-42.
- Kočiš Krutilová V, Bahnsen L, De Graeve D. The out-of-pocket burden of chronic diseases: the cases of Belgian, Czech and German older adults. *BMC Health Serv Res*. 2021;21(1):239. doi:10.1186/s12913-021-06259-w
- Kim H, Kim EJ. Effects of senility on disease burden: panel analysis using 2010 to 2019 OECD health data. *Health Serv Insights*. 2023;16:11786329231188286. doi:10.1177/11786329231188286
- Howard SW, BeLue R, van de Klundert J. Editorial: the role of healthcare delivery, payment and policy innovations in decreasing the global burden of chronic disease. *Front Public Health*. 2022; 10:972064. doi:10.3389/fpubh.2022.972064.
- Dragos SL, Mare C, Dragos CM, Muresan GM, Purcel AA. Does voluntary health insurance improve health and longevity? Evidence from European OECD countries. *Eur J Health Econ*. 2022;23(8):1397-1411. doi:10.1007/s10198-022-01439-9
- Cecchini M, Vuik S, Everard C. Cluster Analysis of Non-communicable Disease Burden, Prevention and Management Across EU27+2 Countries. OECD Health Working Papers No. 195. OECD Publishing; 2026. doi:10.1787/eba334a3-en
- Institute for Health Metrics and Evaluation (IHME). Global Burden of Disease (GBD) Results Tool. IHME. 2025. Accessed October 7, 2026. <https://vizhub.healthdata.org/gbd-results/>
- OECD. OECD Health Statistics. OECD Data Explorer. 2025. Accessed October 7, 2026. [https://data-explorer.oecd.org/?fs\[0\]=Topic%2C0%7CHealth%23HEA%23&pg=0&fc=Topic&bp=true&snb=85](https://data-explorer.oecd.org/?fs[0]=Topic%2C0%7CHealth%23HEA%23&pg=0&fc=Topic&bp=true&snb=85)
- World Bank. World Development Indicators. World Bank Open Data. 2026. Accessed October 7, 2026. <https://databank.worldbank.org/source/world-development-indicators>
- Hennig C. Cluster-wise assessment of cluster stability. *Comput Stat Data Anal*. 2007;52(1):258-271. doi:10.1016/j.csda.2006.11.025
- Stuckler D. Population causes and consequences of leading chronic diseases: a comparative analysis of prevailing explanations. *Milbank Q*. 2008;86(2):273-326. doi:10.1111/j.1468-0009.2008.00522.x
- López-Casasnovas G, Soley-Bori M. The socioeconomic determinants of health: economic growth and health in the OECD countries during the last three decades. *Int J Environ Res Public Health*. 2014;11(1):815-829. doi:10.3390/ijerph110100815
- Alkaya A, Alkaş C. OECD ülkelerinin sağlık göstergelerine göre kümeleme analizi sınıflaması. *Sosyal Güvençe*. 2021;19:427-474. doi:10.21441/sosyalguvence.1050881
- Ekinci R. Kronik hastalıkların sosyoekonomik belirleyicileri üzerine bir makine öğrenmesi analizi. *İzmir İktisat Dergisi*. 2025;40(4):942-967. doi:10.24988/ije.1491030
- WHO Regional Office for Europe. WHO European Regional Obesity Report 2022. WHO Regional Office for Europe; 2022. Accessed May 2, 2026. <https://www.who.int/europe/publications/i/item/9789289057738>
- Yiğit P, Yılmaz E, Gedikli E, Hayran OE. The relationship between noncommunicable diseases risk factors, GINI coefficient and health status in organization of economic development and cooperation countries. *Türkiye Klinikleri J Health Sci*. 2021;6(3):598-608. doi:10.5336/healthsci.2020-77423
- Koene RJ, Prizment AE, Blaas A, Konety SH. Shared risk factors in cardiovascular disease and cancer. *Circulation*. 2016;133(11):1104-1114. doi:10.1161/CIRCULATIONAHA.115.020406
- Peasey A, Bobak M, Kubinova R, et al. Determinants of cardiovascular disease and other non-communicable diseases in Central and Eastern Europe: rationale and design of the HAPIEE study. *BMC Public Health*. 2006;6:255. doi:10.1186/1471-2458-6-255
- Movsisyan NK, Vinciguerra M, Medina-Inojosa JR, Lopez-Jimenez F. Cardiovascular diseases in central and Eastern Europe: a call for more surveillance and evidence-based health promotion. *Ann Glob Health*. 2020;86(1):21. doi:10.5334/aogh.2713
- Cenko E, Manfrini O, Fabin N, et al. Clinical determinants of ischemic heart disease in Eastern Europe. *Lancet Reg Health Eur*. 2023;33:100698. doi:10.1016/j.lanepe.2023.100698
- WHO Regional Office for Europe. Health and Care Workforce in Europe: Time to Act. WHO Regional Office for Europe; 2022. Accessed May 5, 2026. <https://www.who.int/europe/publications/i/item/9789289058339>
- Adolph TE, Tilg H. Western diets and chronic diseases. *Nat Med*. 2024; 30(8):2133-2147. doi:10.1038/s41591-024-03165-6
- Anwar A, Hyder S, Mohamed Nor N, Younis M. Government health expenditures and health outcome nexus: a study on OECD countries. *Front Public Health*. 2023;11:1123759. doi:10.3389/fpubh.2023.1123759

28. Aytemiz L, Sart G, Bayar Y, Danilina M, Sezgin FH. The long-term effect of social, educational, and health expenditures on indicators of life expectancy: an empirical analysis for the OECD countries. *Front Public Health*. 2024;12:1497794. doi:10.3389/fpubh.2024.1497794
29. Masseria C, Mossialos E, Allin S. Measurement of Socioeconomic Inequality of Health in 10 European Countries: An Exploratory Analysis of SHARE Using Three Approaches. European Commission Research Note; 2006.
30. Wang F, Wang JD. Investing preventive care and economic development in ageing societies: empirical evidences from OECD countries. *Health Econ Rev*. 2021;11(1):18. doi:10.1186/s13561-021-00321-3