EAI Endorsed Transactions
on Pervasive Health and Technology

Research Article EAI.EU

Related factors with NCD in developing countries: economic, diet and risk factors dimensions

S. A. Dominguez-Miranda[1,*] and R. Rodriguez-Aguilar[1]

[1]Facultad de Ciencias Económicas y Empresariales, Universidad Panamericana, Augusto Rodin 498, 03920, Mexico City, México

Abstract

INTRODUCTION: Noncommunicable diseases (NCD) such as cardiovascular diseases, cancers, respiratory diseases, and diabetes mellitus are leading causes of global mortality. These conditions are often linked to lifestyle factors such as inadequate physical activity, poor diet, smoking, and excessive alcohol consumption. The economic burden of NCD is substantial, impacting both health systems and the broader economy, and severely affecting individuals' quality of life. Analysing mortality causes and applying mathematical models can help monitor health trends, understand disease patterns, assess intervention effectiveness, and inform health decision-making.

OBJECTIVES: This study aims to evaluate the relationship between economic, dietary, and health risk factors and NCD among the economically active populations in 13 developing countries for 2019. The objectives include applying dimensionality reduction to detect variability across countries, analysing behavioural patterns of key variables, and generating indices to monitor NCD-related factors.

METHODS: A dataset for 2019 was compiled, including 76 variables related to economic, dietary, and lifestyle factors for 13 developing countries. Principal component analysis (PCA) was used to reduce dimensions and group relevant information, focusing on four NCD: cardiovascular diseases, chronic respiratory diseases, neoplasms, and diabetes mellitus. Indices for diet, economic factors, and risk factors related with mortality were created. A clustering model grouped countries based on these indices, and a Random Forest model was employed to identify significant predictors of NCD mortality.

RESULTS: Some relevant characteristics were identified in the countries analyzed, as well as interesting patterns among the factors related to NCD. The countries could be grouped considering their economic and nutritional behavior. It was observed that Latin American countries and Poland behave similarly, just as Asian countries show a similarity in eating behavior. The economic indicators of investment in health, as well as hours worked, behave in a similar way. It was identified that there are certain foods that have a similar behavior both in their consumption and in how they affect NCD. Thanks to the elaboration of the indices, it was observed that the countries of the Middle East and North Africa have a better food balance, but not the countries of Latin America. Using a Random Forest model the cardiovascular influence was identified as the most important predictors, highlighting the critical influence of cardiovascular health factors on NCD mortality.

CONCLUSION: The application of a dimensionality reduction method and cluster analysis out of quantitative methods made it possible to characterize the behavior of a set of variables that impact NCD, as well as to synthesize this information into specific indices by category of analysis. Strategies focused on improving NCD indicators can have a greater impact by identifying similar behavior profiles among developing countries, in the same way, joint policies could be designed to address NCD through specific actions by dimension of analysis and extend these policies to countries with similar profiles. The use of the Random Forest model further underscored the importance of cardiovascular health factors, suggesting that targeted interventions in this area could significantly reduce NCD mortality.

Keywords: Noncommunicable diseases, principal components analysis, cluster analysis, random forest, pervasive healthcare

Received on 29 June 2023, accepted on 27 September 2024, published on 17 October 2024

Copyright © 2024 S. A Dominguez-Miranda *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution license, which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eetpht.10.3499

[*]Corresponding author. Email: 0246533@up.edu.mx

1. Background

For just over 20 years, NCD have occupied the first places as causes of general death: heart disease, stroke, and diabetes *mellitus*, being in first, second and ninth place respectively [1]. Researchers showed that the main reasons for consultation by age group for the 20-49 age group are acute respiratory diseases and symptoms, followed by diabetes, cardiovascular disease, and obesity [2].

A high level of abnormal concentration of cholesterol, triglycerides, HDL cholesterol and uric acid in the first years of adult life can increase the incidence of the disease and years later arterial hypertension may appear [3]. Having a high body mass index greater than 25, hypertension and altered glucose, the well-known metabolic syndrome can be generated, and its presence can lead to a risk of contracting diabetes *mellitus*. This is a chronic metabolic disease characterized by high levels of glucose in the blood, which leads over time to serious damage to the heart, blood vessels, eyes, kidneys, and nerves. The most common being type 2 diabetes, usually acquired in adults, which occurs when the body becomes resistant to insulin or does not produce enough of the hormone [4]. Untreated diabetes leads to multi-organ and systemic lesions, including the heart, kidneys, nerves, and blood vessels, which impairs quality of life and increases the mortality rate caused by diabetes complications such as cardiovascular disease, kidney disease, neoplasms and even Alzheimer's [5].

The term cardiovascular disease [2] refers to diseases of the heart and blood vessels, produced mainly by the accumulation of cholesterol plaques on the artery walls, making them narrow. Other causes of cardiovascular diseases are high levels of uric acid, diabetes, obesity, smoking or stress and can be classified as: coronary heart disease, heart failure, arrhythmias, valvular heart disease, hypertension, stroke, or congenital heart disease [6].

Researchers documented that there is an association between systolic blood pressure (SBP) and the risk of developing a cardiovascular event, with a 2 mm/Hg elevation in SBP associated with a 7% increased risk of mortality from ischemic heart disease and 10% increased risk of stroke mortality [7]. Hypertension is considered an important risk factor for myocardial infarction, ischemic and hemorrhagic stroke, heart failure, chronic kidney disease, cognitive impairment and premature death all considered within cardiovascular diseases [8].

Mortality in chronic-respiratory diseases (CRD) [9] refers to the most common chronic respiratory conditions, which are characterized by persistent inflammation of the airways and bronchial obstruction and according to the World Health Organization [10] is the third cause of death in the world causing more than 3.23 million deaths in 2019 and refer to abnormalities in the airways of the lungs that lead to the limitation of airflow in and out of the lungs can cause the airways to narrow or obstruct, causing the destruction of parts of the lung, mucus that blocks the airways, inflammation and swelling of the lining of the airways. The main cause of CRD is tobacco use and genetic predisposition, but there are also other causes often of a work-related nature such as exposure to dust, gases, steam and even pollution [11].

Finally, with respect to neoplasms or tumors, it is the second cause of death in the world, the most common being lung, prostate, colon, stomach and liver in men and breast, colon, cervical lung, and thyroid cancer in women [12].

Companies are susceptible to suffering the impact generated by these diseases on their workers. Causing absenteeism, presenteeism, poor work performance and loss of productivity. Recent research has shown that employees with low levels of physical activity and sedentary behavior are less productive at work, exhibit high presenteeism, have lower work capacity, and get sick more [13].

Therefore, prevention and health promotion strategies are sought to minimize this impact. Programs implemented in developing countries are being strengthened as part of a commitment to inform the global monitoring framework through multisectoral national action plans for NCD, although they are primarily based in primary care [14].

We can take examples of policies that are in the process of implementation such as the strategy of the NCD prevention and control program by the World Health Organization [15] where four strategic lines are established for the prevention of these diseases in line with 25 indicators and the 9 targets contained in the WHO global framework of comprehensive surveillance which are:

a) Multisectoral policies and partnerships for NCD prevention and control: Strengthen and promote multisectoral actions with all relevant sectors of government and society, including integration into economic, academic and development agendas.

b) NCD risk factors and protective factors: Reduce the prevalence of major NCD risk factors and strengthen protective factors; employ evidence-based health promotion strategies and policy instruments, including regulation, surveillance, and voluntary measures; and addressing the social, economic, and environmental determinants of health.

c) Health systems response to NCD and their risk factors: Improving coverage, equitable access, and quality of care for the four major NCD (cardiovascular disease, cancer, diabetes, and chronic respiratory diseases) and others that are nationally prioritized, with an emphasis on primary health care that includes prevention and improved self-management.

d) NCD surveillance and research: Strengthen countries' capacity for surveillance and research on NCD, their risk factors and determinants, and use research findings as a basis for evidence-based policy development and implementation, academic, development and implementation programs.

However, studies evaluating the effectiveness of programs to provide primary prevention interventions for the prevention and control of NCD in middle-income country settings are insufficient and under-analysed. Furthermore, the limited results available come mainly from observational studies, and evidence from controlled trials needs to be synthesized [16].

Some researchers used panel data to associate physical inactivity with many respiratory diseases using multi-disciplinary methods [17]. In other research was developed a characteristic indicator framework to influence policy implementation, including democracy, corporate penetration which is an indicator of corporate influence, NCD burden, and prevalence of risk factors using multivariate regression models [18]. Additionally, other researchers sought to relate socioeconomic factors in morbidity in China based on panel data analysis for people over 50 years of age [19]. Also, models based on decision trees and a possible predictability of NCD have been found developed [20] and other various indices had been created for understand the socioeconomic inequalities in NCD for low-middle income countries analyzing four a set of related variables [20, 21]. Several studies have been performed but focused on general population. More variables should be considered to understand better the characterization of NCD in developing countries and mainly for the time where the population is actively working having in mind the workers could be more productive when health indicators are controlled and could impact in the economic growth [22, 23, 24, 25].

In recent years, investigations into the application of principal component analysis for built monitoring indicators have been conducted with varying approaches. In India, a study was conducted to assess the dietary quality and the high prevalence of NCD among participants aged 35 to 55. This study unveiled a negative correlation with meat consumption, in contrast to a positive correlation observed with legume consumption. Furthermore, it revealed a deficiency in vegetable consumption, dietary fiber, and omega-3 fatty acids compared to the World Health Organization (WHO) recommendations [26]. In the same geographical context, disparities in food access and health inequities were analyzed to elucidate NCD risk factors. Factors such as education, dietary choices, and income were scrutinized as contributors to these complex dynamics [27]. In Iran, an investigation centered around the development of an inflammatory diet index aimed to comprehend anthropometric patterns among children and adolescents. Results exhibited a positive association between higher quartiles of the inflammatory diet index and elevated Body Mass Index (BMI) [28]. In the Philippines, an analysis employing data from the 2013 health survey sought to elucidate the connection between dietary quality and the prevalence of cardio-metabolic diseases. Findings demonstrated that lower dietary quality was associated with a higher prevalence of hypertension, dyslipidemia, and obesity [29]. Researchers in Brazil analyzed BMI trends [30]. In South Africa, an endeavor was made to elucidate the impact of tobacco consumption on the adult population and its contribution to the proliferation of NCD, as well as its association with tuberculosis [31]. Additionally, mathematical algorithms have been employed to investigate the socio-economic disparities and their influence on various NCD [32, 33, 34, 35]. In the realm of scientific inquiry, these investigations collectively contribute to the comprehensive understanding of the intricate dynamics governing dietary choices, health outcomes, and socio-economic factors, akin to pieces of a complex puzzle gradually falling into place with the support of mathematical methods.

The main objective of this work is the construction of indices that synthesize the information of a set of variables linked to NCD using principal component analysis, which is a dimensionality reduction technique that aims to find a new set of uncorrelated variables, that capture most of the variation in the original data [36]. Sets of 76 variables that address different dimensions of the problem were identified and consolidated in 3 indexes that allows monitoring the performance of these factors related to NCD in developing countries. Subsequently, these indices will be used as variables for the estimation of a Random Forest model and to evaluate the importance of the constructed indicators in the mortality of NCD. Finally, a cluster analysis was carried out, with the purpose of synthesizing the economic information of the indices in groups of countries with similar characteristics regarding the factors related to NCD.

2. Methodology

All dietary variables and risk factors contained in the databases used were considered, as well as economic variables selected by the authors focused on NCD as shown in Anex 1. Thirteen developing countries were selected by a non-random method with samples from Europe, Africa, Americas, and Asia shown in the world database from world data in accordance with the United Nations which were: Poland, Turkey, Ukraine, Egypt, Morocco, South Africa, Mexico, Brazil, Chile, Colombia, Argentina, Malaysia, and Thailand [37]. It was considered a filter for the age group between 20 and 64 years because it is considered as the working age.

A principal component analysis (PCA) was performed for generating 3 different indexes considering only the values shown for 2019 to analyze the behavior of the different variables that affect the mortality of the NCD: Cardiovascular diseases (CAD), Chronic-respiratory diseases (Resp), Diabetes *mellitus* (DM) and Neoplasm diseases (Neo) with respect to the behavior in the thirteen selected countries. The first index was focused on economic variables to understand the economic behavior of the selected countries based on the information obtained from the Organization for Economic Co-operation and Development (OECD) [38], World Bank Open Data (WBOD) [39] and American Economic Review (AER) [40]. It was selected 2019 due to analyze factors previous COVID impact.

A second index was built based on the food consumption according with the information obtained from the Food and Agriculture Organization of the United Nations (FAO) [41] with the objective of analyze the food balance along the selected countries. Finally, different variables were combined from the Institute for Health Metrics and Evaluation (IHME) [42], this database contains information not only on the mortality of NCD, but also a registry of the causes of mortality of the diseases with two different levels, the first one when there is a high incidence of the variable, for example, high consumption of sugary drinks or exposure to pollution, these variables were labeled as H-factor. The second level shows information on those variables that, because of the low usability or low administration, affect mortality, for example, low physical activity or low cereal consumption, labeled as L-factor. In the case of chronic-respiratory diseases, only high-cut-off variables, labeled as well as H-factors, that affect the disease were found.

The construction of indices that synthesize information from the dimensions considered will allow for the estimation of subsequent models such as Random Forest to evaluate the importance of each dimension, and Cluster Analysis to group and characterize the selected sample of developing countries (Figure 1).

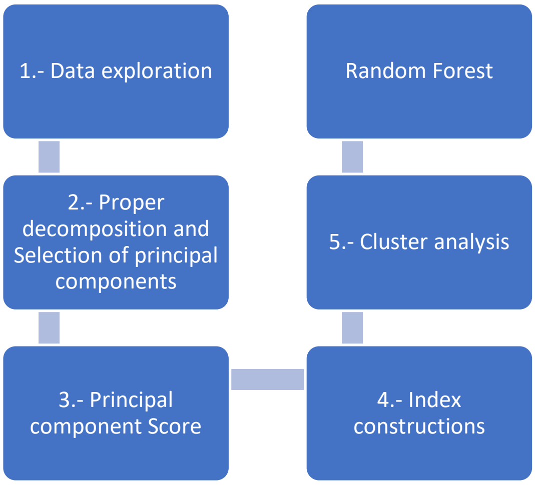

1.- Data exploration
2.- Proper decomposition and Selection of principal components
3.- Principal component Score
4.- Index constructions
5.- Cluster analysis
Random Forest

Figure 1. Cuantitative Modeling Process

Principal component scores are calculated for each observation in the dataset. These scores represent the coordinates of the observations in the small-dimensional space defined by the selected principal components. The weights of the principal components are determined based on their relative importance in capturing the underlying variation in the data. Once the weights of the principal components are determined, a composite index is calculated by aggregating the principal component scores for each observation, weighted by the corresponding weights as reflected in equation 1.

$$Y_i = \frac{x_{1i}w_1 + x_{2i}w_2 + \cdots + x_{ki}w_k}{w_1 + w_2 + \cdots + w_k} \qquad (1)$$

Where Y_i is the value of the index for each set of variables i, x_{ki} is the value of the k factor coordinate selected, and w_i is the proportion of the total inertia explained by that coordinate. Finally, the value of Y_i is adjusted to a range between [0,100] to facilitate its interpretation. The composite index provides a summary measure derived from the original variables, reflecting the underlying patterns and relationships captured by the PCA analysis [43].

Cluster Analysis

Cluster analysis was performed to show the usefulness of the indices and to group countries with similar characteristics regarding factors related to NCD. The application of the Random Forest model will complement the descriptive characterization that could be carried out once the clusters have been estimated. Both elements will allow for a profile per cluster as well as knowing which dimensions represented by the indices are determinants in NCD mortality. This will allow for the identification of lines of action consistent with the profiles of similar developing countries, as well as being able to identify successful good practices implemented in similar countries.

Random Forest

The Random Forest model was estimated considering NCD mortality as the dependent variable and the explanatory variables were the indices constructed with the dimensions of the risk factors considered. The Random Forest model was optimized using a k-fold cross-validation criterion and a performance metric based on the RMSE. Additionally, a variable importance estimation method was applied to rank the relevance of the indices in determining NCD mortality. To do this, the variable importance method by permutation was used, which calculates the importance of the variable by evaluating the changes in the performance of the model when the values of a variable are randomly shuffled.

3. Results

3.1 Index Estimated

The indexes generated from the PCA methodology focused on the behavior of economic variables, food balance and risk factors influenced by dietary and lifestyle on NCD mortality. The number of principal components selected for each group of variables as well as the total variance explained are shown in Table 1.

The selection of the number principal components seeks capturing on average at least 80% of the total variability of each group of variables, and also allows building two-dimensional graphs to characterize the countries analysed in the sample.

Table 1. PCA cumulative portion

Name	Index 1 Economic	Index 2 Food Balance	Index 3 Risk factors			
			CAD	Resp	DM	Neo
Dimensions selected	2	4	2	2	2	2
Variance cumulative proportion	86.84%	78.06%	90%	94%	91%	88%

Health Economic Index

Figure 2 shows the graph of the first two principal components and their correlation with each original variable with the principal component variables. Additionally, the location of each country in the sample in this two-dimensional space allows us to locate the position of the country with respect to each principal component.

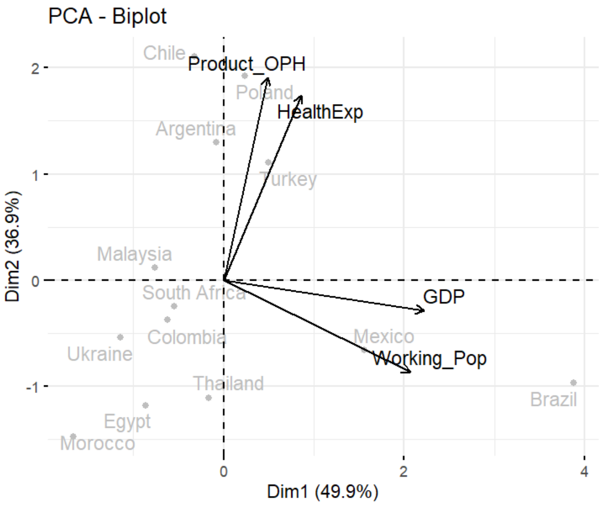


Figure 2. Correlation between economic variables and countries in the first two dimensions.

The second component groups together GDP and the working-age population, while the first component groups together information on health expenditure and hours worked. On the other hand, it is observed that the countries of Malaysia, Turkey, Thailand, Colombia, Egypt, Ukraine, South Africa, and Morocco are like the selected economic indicators. Poland, Chile, Turkey, and Argentina have similar behavior and, Mexico and Brazil show close performances.

Table 2 shows the generated economic index that allows the synthesis of the selected economic variables linked to NCD.

This index allows the ranking of countries based on the performance of the selected economic variables. Brazil, Poland, Turkey and Chile are in the first places of the economic index, above Mexico. In the last places of the ranking are Ukraine, Egypt and Morocco.

Table 2. Economic Index

Position	Country	Economic Index
1	Brazil	100
2	Poland	74.56322
3	Turkey	68.85238
4	Chile	67.48465
5	Mexico	64.64328
6	Argentina	61.42353
7	Malaysia	35.23268
8	South Africa	34.25274
9	Colombia	31.34772
10	Thailand	29.98927
11	Ukraine	20.50134
12	Egypt	17.12691
13	Morocco	0

3.2 Food consumption

The second set of variables to be analyzed corresponds to the food consumption of the selected countries based on calories. The objective is to find patterns regarding food consumption among the selected countries. It can be seen in Figure 3 where the graph of the first two dimensions captures 47% of the total variability.

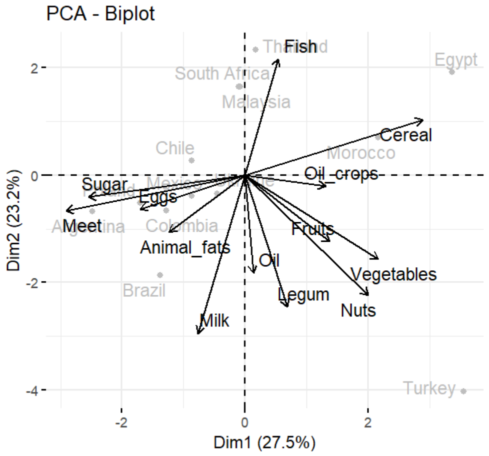


Figure 3. Correlation between dietary variables in the first two dimensions

In terms of diet, sugar, meat, eggs, animal fats and milk are found in a similar section (1). Fruits, vegetables, legumes and nuts are grouped in a similar section (2). Cereals, oilseeds and fish are grouped in a section (3). What stands out in this grouping is the finding in the literature that indicates that there are foods that can have a greater impact on the evolution of mortality from the different NCD, and it is precisely the first group of foods that presents these conditions. Groupings of countries are also observed for each group of foods.

Mexico, Chile, Poland, Argentina, Colombia, Brazil, and Ukraine are localized in the section 1. South Africa, Malaysia, Thailand, Egypt, and Morocco in section 3, which confirms the dietary trend in these countries that is more plant-based than meat-based. Turkey is the only one that is isolated and more section 2.

The construction of the food index considers the consumption of all food groups by country, this means that it is possible to rank the country that consumes the most foods. However, one food group is considered healthy food and another unhealthy food. Therefore, a segmented index was built, it was sectioned food consumption, considering the groups found in Figure 3, sugar, eggs, meat, animal fats and milk were grouped into Group 1 and the rest of the food into Group 2. With the intention of better understanding food consumption but based on those foods that have a higher caloric intake and that is corroborated with figure 3 of those foods that, according to their literature, could have had benefits in the diet and that are found on the positive side of dimension 1 of Figure 3. This is how the results are observed in Table 3.

Table 3. Segmented food consumption Index

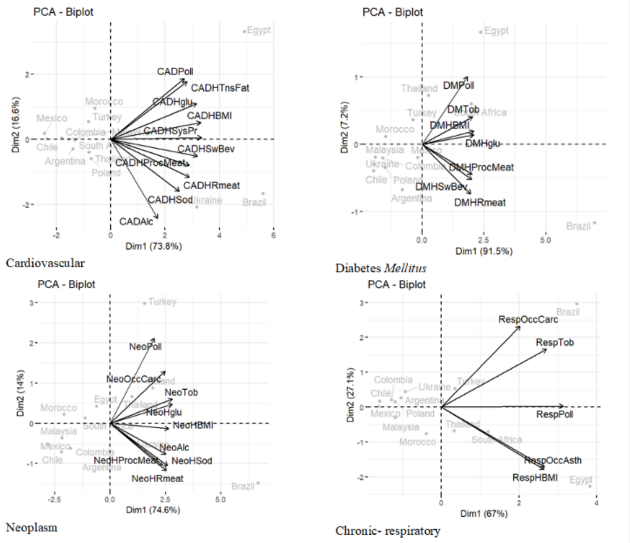
Country	Group 1	Group 2
Argentina	100.00	1.28
Brazil	78.89	19.91
Chile	66.17	17.62
Colombia	86.17	10.88
Egypt	-	51.96
Malaysia	64.00	40.37
Mexico	89.92	23.71
Morocco	24.73	41.70
Poland	82.22	13.81
South Africa	28.60	-
Thailand	51.36	26.40
Turkey	38.13	100.00
Ukraine	73.11	18.09

An interesting contrast can be observed between a group of countries in terms of the food group they consume in the highest proportion. In the case of Latin American countries, together with Poland, a higher consumption of Group 1 is observed. The case of Egypt is interesting, where it is observed that they have a high consumption of foods from Group 2, contrary to Brazil or Argentina, whose food consumption index is more focused on Group 1. In the case of Mexico, its consumption is concentrated on foods from Group 1 compared to Group 2.

Risk Factors Index

H-Factors

Those variables that appear to have a high influence on mortality and that, due to their characteristics, appear more frequently in the statistics of the databases, were called Factor H. Figure 4 shows how the causes of death are grouped for each NCD, as well as the grouping of countries according to the first two principal components. In the case of cardiovascular diseases, it was observed that the BMI behaves in a similar way with high levels of glucose, trans fats or hypertension, which are indicators of health. In the case of Diabetes Mellitus and neoplastic diseases, health style behaves in a similar way as does the BMI, high levels of glucose and in this case tobacco consumption and pollution influence. It is observed that in chronic respiratory diseases, tobacco and exposure to carcinogens are grouped in a similar way. In all cases, high consumption of foods such as processed meat, sugary drinks or alcohol consumption are grouped together.



Cardiovascular

Diabetes *Mellitus*

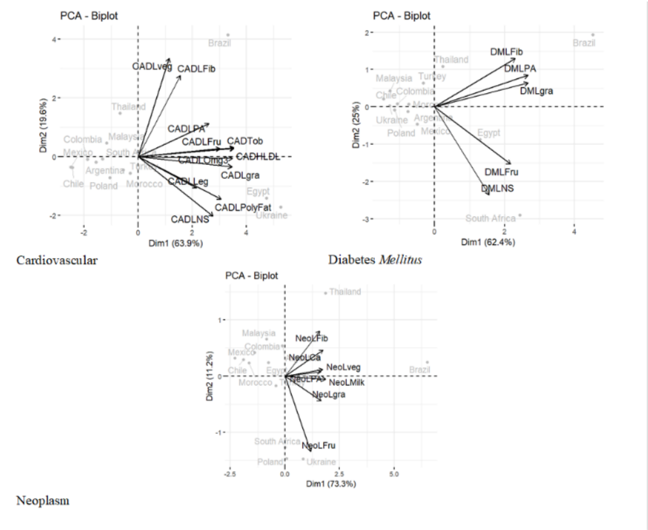
Neoplasm

Chronic- respiratory

Figure 4. Correlation between H-Factors and countries in the first two dimensions by NCD.

L-Factors

An evaluation was carried out on the four selected NCD, those variables that appear as a low percentage of their use that affect mortality and that, due to their characteristics, appear less frequently in the statistics of the databases, were called L Factors, which are shown in Figure 5.



Cardiovascular

Diabetes *Mellitus*

Neoplasm

Figure 5. Correlation between L-Factors and countries in the first two dimensions by NCD.

When analyzing the correlation of countries and L Factors in the first two dimensions generated by the principal components, it is observed how the countries are positioned with respect to the risk factors, a similar behavior stands out in almost all countries, except Brazil and Egypt.

Table 4 shows the ranking of countries generated with the H and L Factor indices by NCD. It is clarified that the position in the H indices refers to the country with the greatest impact of the H values, that is, those countries that have been impacted by the high frequency of the variables, for example, high glycemic index or high prevalence of hypertension. Likewise, the list of countries influenced by the L factors are those that show a behavior of not using certain variables in favor of the mitigation of NCD, for example, low physical activity or low consumption of omega 3.

Table 4. Risk factor (H and L) index by NCD

Country	Cardiovascular Index H	Cardiovascular Index L	Diabetes Mellitus Index H	Diabetes Mellitus Index L	Neoplasm Index H	Neoplasm Index L	Respiratory Index H
Argentina	11.60	13.56	13.13	8.37	29.52	25.16	10.91
Brazil	94.56	97.95	100.00	100.00	100.00	100.00	100.00
Chile	0.74	0.75	-	-	6.13	4.44	1.50
Colombia	7.53	9.52	4.16	5.18	10.72	10.54	6.60
Egypt	100.00	94.24	48.42	34.84	25.60	17.44	71.77
Malaysia	10.52	21.44	0.78	4.23	7.17	17.15	4.91
Mexico	-	-	19.96	10.42	-	-	-
Morocco	25.25	29.00	5.57	9.89	9.34	7.21	14.70
Poland	19.23	18.21	5.29	3.06	53.70	23.96	7.80
South Africa	13.32	17.30	44.19	39.69	19.33	23.47	40.25
Thailand	18.80	32.45	24.85	30.61	43.13	49.30	26.48
Turkey	21.25	25.96	17.92	19.68	53.90	20.48	34.79
Ukraine	63.32	100.00	0.30	1.51	46.74	32.44	13.76

It can be observed that Brazil is among the top indices with the greatest impact on NCD. On the other hand, it can be observed that Mexico, with respect to cardiovascular, oncological and chronic respiratory diseases, is the country with the best performance compared to other developing countries, but not with respect to diabetes mellitus, since it is in the middle of the generated index.

The previous analysis allows us to identify potential relationships between the indices and mortality, as well as a natural grouping between selected countries. The following sections show the results generated by the Random Forest and Cluster models, which seek to identify the effect of each index on mortality, as well as to categorize groups of countries according to the behavior of the generated indices.

3.2 Cluster analysis

A cluster analysis was performed to find groupings of countries according to the developed indices. As a first approximation, a two-dimensional graph was created comparing the health economic index and the two food consumption indices by group, in relation to mortality from the four selected NCD. Figure 6 shows that there are natural groupings of countries according to the behavior of the indices vs. mortality.

Analysing the health economic index, it can be seen that countries such as Ukraine or Egypt, which are below the economic index, have a higher mortality rate; other countries show that being at intermediate values of the index, as is the case of Chile, Argentina, Poland or Turkey, the mortality indicator is lower than the other countries; in the case of Mexico or Brazil, although they show a higher economic index, the results have not been favourable with respect to mortality due to NCD.

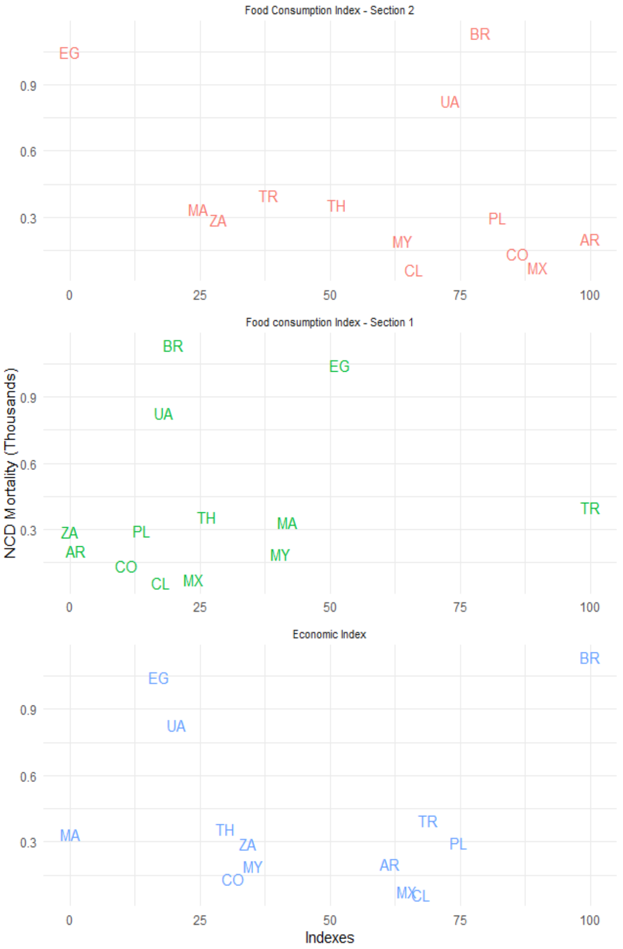


Figure 7. Indexes vs NCD Mortality

On the other hand, when analyzing the food consumption indexes by group, it is observed that countries such as Mexico, Ukraine or Brazil, which show a higher number within the index of food consumption that were considered within Group 1, which are sugar, eggs, meat, animal fats and milk, have a moderate to high mortality rate. Countries such as Malaysia, South Africa or Turkey, which are below the index, have lower mortality in NCD. However, a different behavior is observed in countries such as Argentina, which being number 1 in the index, does not present a significant mortality rate, or Egypt, which although it is in the lower part of the index, shows a moderately high pattern. Continuing with the consumption of foods in Group 2, Turkey, which is number 1 in the index due to the highest consumption of these foods, shows a low mortality rate, in contrast to countries such as Brazil or Ukraine, which were below this index, present a higher mortality rate.

It is worth mentioning that the use of the indices shown represents an exploratory analysis of the information constructed. In the case of groupings of countries by profile based on the generated indices, a cluster analysis was estimated using the K-means method (Fig. 8).

Three partitions were considered: group 1 includes 7 countries, group 2 6 countries, and group 3 would be formed only by Brazil. Group 1 presents a healthier average profile in terms of diet, economic index, and risk factors. Group 2 represents those countries with a lower economic index, and a higher diet concentrated in Group 1 of unhealthy foods, as well as higher risk factors related to NCDs. It is important to highlight that the K-means method integrates information from all the indices simultaneously, which allows the constructed groups to show us a profile of similar countries. This will allow the subsequent analysis of successful public policies adopted and their potential replication within the analyzed cluster.

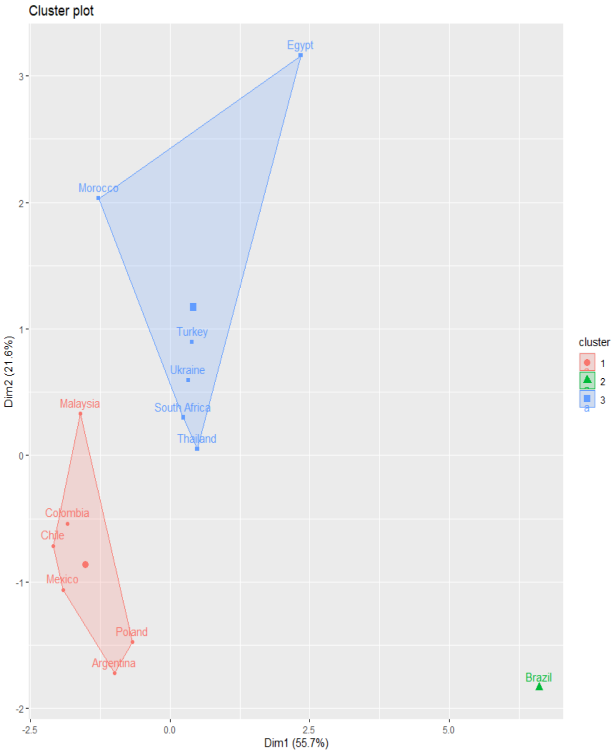


Figure 8. K-means indexes clustering

3.3 Importance of variables (Random Forest)

To determine the importance of the index in NCD mortality, a Random Forest model was estimated, and the analysis of the importance was generated to rank the variables by relevance using all indexes generated, after recalibrating the model with 500 trees, a maximum depth of 5, and a permutation-based importance metric, the prediction accuracy for NCD mortality was evaluated obtaining the RMSE improved to 23,772.16 from the original models of 28,296.42. The importance of each predictor was assessed and is shown in Figure 9. The predictors with the highest importance scores were the Factor H Index and the Factor L Index for cardiovascular diseases.

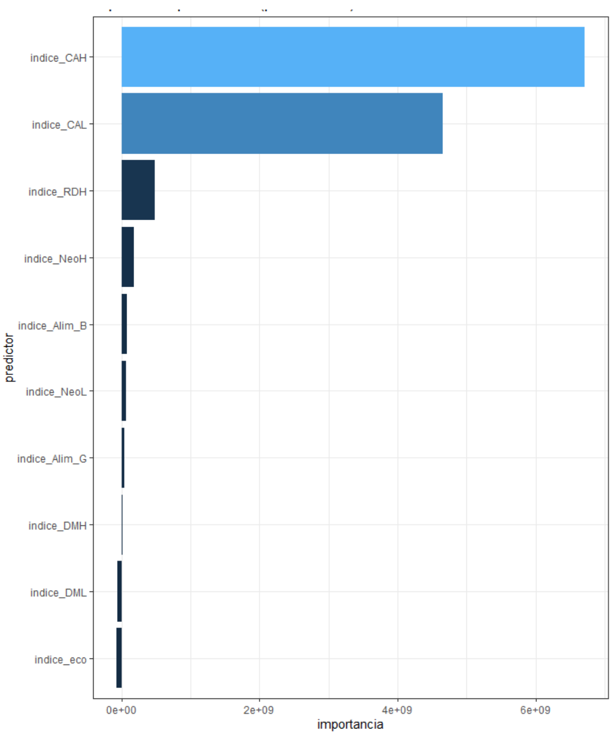


Figure 9. Predictor´s importance – Permutation method

These were followed by the Factor H Index - Chronic Respiratory and the Factor H Index - Neoplasia. The health economic index is ranked fifth in importance, the cluster to which each country belongs is ranked sixth. The food indices of Group 1 and 2 are ranked 7th and 8th, respectively.

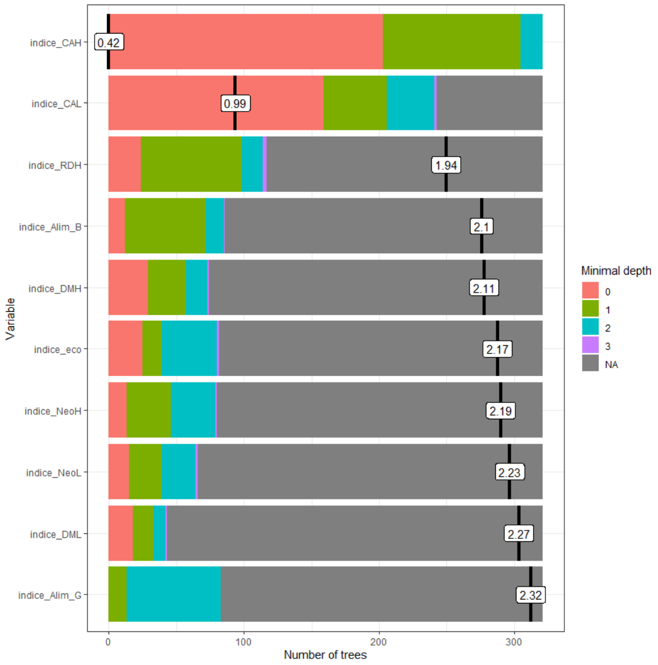


Figure 10: Distribution of minimal depth and its mean

Additionally, the distribution of the minimum depth for each variable was analyzed, revealing that the Cardiovascular Factor H Index frequently appears at the root level of the trees, underlining its critical role in the model. Other important variables such as the Cardiovascular Factor L Index, the Chronic Respiratory Factor H Index and the cluster also showed consistent appearances at shallower depths, contributing significantly to the prediction of NCD mortality (Fig. 9).

These findings including the distribution of minimal depth (Fig.10) highlight the dominant influence of socioeconomic and health indices on NCD mortality in the countries analyzed. The visualization of the importance of the variables confirms the prominent role of the H-factor index - cardiovascular and the L-factor index - cardiovascular, suggesting that targeted policy interventions could potentially mitigate the impact of NCD.

4. Discussions

The presented study allowed to integrate information from a set of variables in the economic, dietary and NCD risk factor dimensions, showing similarities in these dimensions among the developing countries analysed. This similarity in patterns, particularly in health expenditure and food consumption, is fundamental to understanding the factors influencing NCD mortality. The indices generated through PCA provided a comprehensive view of the underlying variables affecting NCD mortality, offering a robust framework for analysis. It is interesting to note how certain dietary factors naturally cluster together due to similar behaviours. These clusters often include foods that have been identified in the literature as having a positive or negative impact on health.

Economic factors such as GDP and health expenditure were found to be significantly correlated with health outcomes, as illustrated through principal component analysis. This relationship underlines the importance of considering economic and occupational factors when addressing health outcomes.

Cluster analysis revealed that Latin American countries, along with Poland, tend to behave similarly, while European, Asian and African countries form another distinct group. This grouping highlights how economic indicators are intricately linked with NCD risk factors. Both the construction of the indices and the cluster analysis successfully illustrated the behaviour of developing countries, highlighting how some can be grouped by their homogeneity. However, differences were found in countries such as Brazil, which, despite being a developing country, does not necessarily display behaviours like other countries in this category. This finding suggests the need for tailored strategies that take such disparities into account.

The Random Forest model provided additional insight into the predictors of NCD mortality. The accuracy of the model, along with its ability to handle many variables and interactions, made it a strong choice for this analysis. The importance of the health economic index and the food indices were corroborated. In particular, the Random Forest model identified the H-factor index and the L-factor index as significant predictors of CVD. The visualization of variable importance and the minimum depth analysis confirm the critical role of specific health indices, particularly those related to cardiovascular health. This knowledge suggests that targeted policy interventions targeting these key predictors could potentially mitigate the impact of NCD in developing countries. Likewise, the variable importance ranking will allow prioritizing interventions based on the relevance of the factor related to NCD.

5. Conclusions

The study highlighted the significant relationship between economic, dietary and health factors and NCD mortality in developing countries. The use of principal component analysis in selected countries helps to understand the behavior of the underlying variables. The generation of indices allows for timely monitoring of factors related to NCD mortality. Cluster analysis provides a deeper insight into the profiles of the groups of countries formed and represents a valuable input for the analysis of health policies successfully implemented in these clusters, allowing successful practices to be adopted by the rest of the countries in the cluster. Random Forest analysis improved the understanding of the most influential predictors of NCD mortality. The high accuracy of the model underlines the reliability of the findings and supports the development of specific interventions. The identification of significant predictors such as the H-factor index and the L-factor index for cardiovascular diseases suggests that these factors should be prioritized in policy formulation and health initiatives.

The study suggests that strategies aimed at improving NCD indicators could be more effective in countries that show similar behaviors. Shared sectoral programs could benefit the working class by addressing common health challenges observed across the indices. Countries that cluster in terms of economic and dietary behaviors could benefit from collaborative health initiatives tailored to their specific needs. The significant predictors identified by the Random Forest model highlight the need for targeted policy interventions that address the most influential risk factors. These interventions could potentially mitigate the impact of NCD by addressing the most significant determinants.

Overall, the study highlights the importance of multifaceted approaches to address NCD mortality, considering economic, dietary and lifestyle factors. The evidence-based insights provided by this analysis can inform policy makers and health professionals to design effective interventions to reduce NCD mortality and improve population health.

References

[1] WHO, World Health Organization (2014). Global Status Report on noncommunicable diseases 2014. Last accessed 06/15/2023. https://apps.who.int/iris/handle/10665/148114

[2] Shamah-Levy, T., Vielma-Orozco, E., Heredia-Hernández, O., Romero-Martínez, M., Mojica-Cuevas, J., Cuevas-Nasu, L., Santaella-Castell, J.A. & Rivera-Dommarco, J. (2020). *National Health and Nutrition Survey 2018-19: National Results.* National Institute of Public Health.

[3] Aguilar, C.A. (1999). Health promotion for the prevention of chronic-degenerative diseases linked to diet and lifestyle. *Community Health and Health Promotion*. ICEPSS Publishers.

[4] WHO, World Health Organization. (2022). Diabetes https://www.who.int/health-topics/diabetes#tab=tab_1

[5] Li, S., Wang, J., Zhang, B., Li, X., & Liu, Y. (2019). Diabetes mellitus and cause-specific mortality: a population-based study. *Diabetes & metabolism journal*, *43*(3), 319-341. https://doi.org/10.4093/dmj.2018.0060

[6] WHO, World Health Organization. (2021). Cardiovascular diseases (CVDs). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[7] Mancia, G., Fagard, R., Narkiewicz, K., Redán, J., Zanchetti, A., Böhm, M., ... & Zannad, F. (2013). 2013 Practice guidelines for the management of arterial hypertension of the European Society of Hypertension (ESH) and the European Society of Cardiology (ESC): ESH/ESC Task Force for the Management of Arterial Hypertension. *Journal of hypertension, 31*(10), 1925-1938. https://doi.org/10.3109/08037051.2013.817814

[8] NICE. (2022). Clinical guideline. Hypertension in adults: diagnosis and management: National Institute for Health and Care Excellence (NICE). https://www.nice.org.uk/guidance/ng136/resources/hypertension-in-adults-diagnosis-and-management-pdf-66141722710213

[9] Scichilone, N., Benfante, A., Bocchino, M., Braido, F., Paggiaro, P., Papi, A., ... & Sanduzzi, A. (2015). Which factors affect the choice of the inhaler in chronic obstructive respiratory diseases? *Pulmonary Pharmacology & Therapeutics*, *31*, 63-67. https://doi.org/10.1016/j.pupt.2015.02.006

[10] WHO, World Health Organization. (2022). Chronic obstructive pulmonary disease (COPD). https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd)

[11] Mannino, D. M., Watt, G., Hole, D., Gillis, C., Hart, C., McConnachie, A., ... & Vestbo, J. (2006). The natural history of chronic obstructive pulmonary disease. *European Respiratory Journal*, *27*(3), 627-643. https://doi.org/10.1183/09031936.06.00024605

[12] WHO, World Health Organization. (2022). Cancer. https://www.who.int/health-topics/cancer#tab=tab_1

[13] Rongen, A., Robroek, S. J., van Lenthe, F. J., & Burdorf, A. (2013). Workplace health promotion: a meta-analysis of effectiveness. *American journal of preventive medicine*, *44*(4), 406-415. https://doi.org/10.1016/j.amepre.2012.12.007

[14] WHO, World Health Organization (2014). Global Status Report on noncommunicable diseases 2014 Last accessed 06/15/2023. https://apps.who.int/iris/handle/10665/148114

[15] WHO, World Health Organization. (2013). Plan of Action for the Prevention and Control of Noncommunicable

Diseases in the Americas 2013-2019. https://www.paho.org/hq/dmdocuments/2015/plan-accion-prevencion-control-ent-americas.pdf

[16] Jeet, G., Thakur, J. S., Prinja, S., & Singh, M. (2017). Community health workers for non-communicable diseases prevention and control in developing countries: evidence and implications. *PloS one, 12*(7), e0180640. https://doi.org/10.1371/journal.pone.0180640

[17] Ding, D., Lawson, K. D., Kolbe-Alexander, T. L., Finkelstein, E. A., Katzmarzyk, P. T., Van Mechelen, W., & Pratt, M. (2016). The economic burden of physical inactivity: a global analysis of major non-communicable diseases. *The Lancet, 388*(10051), 1311-1324. https://doi.org/10.1016/S0140-6736(16)30383-X

[18] Allen, L. N., Wigley, S., & Holmer, H. (2021). Implementation of non-communicable disease policies from 2015 to 2020: a geopolitical analysis of 194 countries. *The Lancet Global Health, 9*(11), e1528-e1538. https://doi.org/10.1016/S2214-109X(21)00359-4

[19] Zhao, Y., Atun, R., Oldenburg, B., McPake, B., Tang, S., Mercer, S. W., ... & Lee, J. T. (2020). Physical multimorbidity, health service use, and catastrophic health expenditure by socioeconomic groups in China: an analysis of population-based panel data. *The Lancet Global Health, 8*(6), e840-e849. https://doi.org/10.1016/S2214-109X(20)30127-3

[20] Wang, Y., & Wang, J. (2020). Modelling and prediction of global non-communicable diseases. *BMC public health, 20*(1-13). https://doi.org/10.1186/s12889-020-08890-4

[21] Hosseinpoor, A. R., Bergen, N., Kunst, A., Harper, S., Guthold, R., Rekve, D., ... & Chatterji, S. (2012). Socioeconomic inequalities in risk factors for non communicable diseases in low-income and middle-income countries: results from the World Health Survey. *BMC public Health, 12*(1), 1-13. https://doi.org/10.1186/1471-2458-12-912

[22] Shikdar, A. A., & Sawaqed, N. M. (2003). Worker productivity, and occupational health and safety issues in selected industries. *Computers & industrial engineering, 45*(4), 563-572. https://doi.org/10.1016/S0360-8352(03)00074-3

[23] Kirsten, W. (2008). Health and productivity management in Europe. *International journal of workplace health management, 1*(2), 136-144. https://doi.org/10.1108/17538350810893928

[24] Saha, S. (2013). Impact of health on productivity growth in India. *Inter J Eco, Finance & Manag, 2*(4). https://www.ejournalofbusiness.org/archive/vol2no4/vol2no4_6.pdf

[25] Siddique, H. M. A., Mohey-ud-din, G., & Kiani, A. (2020). Human health and worker productivity: evidence from middle-income countries. *International Journal of Innovation, Creativity and Change, 14*(11), 523-544. Available at SSRN: https://ssrn.com/abstract=3748998

[26] Naicker, A., Venter, C. S., MacIntyre, U. E., & Ellis, S. (2015). Dietary quality and patterns and non-communicable disease risk of an Indian community in KwaZulu-Natal, South Africa. Journal of Health, Population and Nutrition, 33, 1-9. https://doi.org/10.1186/s41043-015-0013-1

[27] Pomeroy-Stevens, A., Bachani, D., Sreedhara, M., Boos, J., Amarchand, R., & Krishnan, A. (2022). Exploring urban health inequities: the example of non-communicable disease prevention in Indore, India. Cities & health, 6(4), 726-737. https://doi.org/10.1080/23748834.2020.1848327

[28] Aslani, Z., Qorbani, M., Hébert, J. R., Shivappa, N., Motlagh, M. E., Asayesh, H., ... & Kelishadi, R. (2019). Association of Dietary Inflammatory Index with anthropometric indices in children and adolescents: the weight disorder survey of the Childhood and Adolescence Surveillance and Prevention of Adult Non-communicable Disease (CASPIAN)-IV study. British Journal of Nutrition, 121(3), 340-350. https://doi.org/10.1017/S0007114518003240

[29] Angeles-Agdeppa, I., Sun, Y., & Tanda, K. V. (2020). Dietary pattern and nutrient intakes in association with non-communicable disease risk factors among Filipino adults: A cross-sectional study. Nutrition journal, 19(1), 1-13. https://doi.org/10.1186/s12937-020-00597-x

[30] Felisbino-Mendes, M. S., Cousin, E., Malta, D. C., Machado, Í. E., Ribeiro, A. L. P., Duncan, B. B., ... & Velasquez-Melendez, G. (2020). The burden of non-communicable diseases attributable to high BMI in Brazil, 1990–2017: Findings from the Global Burden of Disease Study. Population Health Metrics, 18(1), 1-13. https://doi.org/10.1186/s12963-020-00219-y

[31] Chidumwa, G., Olivier, S., Ngubane, H., Zulu, T., Sewpaul, R., Kruse, G., ... & Wong, E. B. (2023). Tobacco smoking and prevalence of communicable and non-communicable diseases in rural South Africa: A cross-sectional study. https://doi.org/10.21203/rs.3.rs-2730894/v1.

[32] Gatimu, S. M., & John, T. W. (2020). Socioeconomic inequalities in hypertension in Kenya: a decomposition analysis of 2015 Kenya STEPwise survey on non-communicable diseases risk factors. International journal for equity in health, 19, 1-11. https://doi.org/10.1186/s12939-020-01321-1.

[33] Zere, E., Mandlhate, C., Mbeeli, T. et al. Equity in health care in Namibia: developing a needs-based resource allocation formula using principal components analysis. Int J Equity Health 6, 3 (2007). https://doi.org/10.1186/1475-9276-6-3.

[34] Ochola, S., Kanerva, N., Wachira, L. J., Owino, G. E., Anono, E. L., Walsh, H. M., ... & Fogelholm, M. (2023). Wealth and obesity in pre-adolescents and their guardians: A first step in explaining non-communicable disease-related behaviour in two areas of Nairobi City County. PLOS Global Public Health, 3(2). https://doi.org/10.1371/journal.pgph.0000331.

[35] Liu, L., Wu, X., Li, H. F., Zhao, Y., Li, G. H., Cui, W. L., ... & Cai, L. (2023). Trends in the Prevalence of Chronic Non-Communicable Diseases and Multimorbidity across Socioeconomic Gradients in Rural Southwest China. The journal of nutrition, health & aging, 1-6. https://doi.org/10.1007/s12603-023-1932-y

[36] Jolliffe, I. (2005). Principal component analysis. *Encyclopedia of statistics in behavioral science*. https://doi.org/10.1002/0470013192.bsa501

[37] World data. Developing countries. Last accessed 06/23/2023. https://www.worlddata.info/developing-countries.php.

[38] OECD, Organization for Economic Cooperation and Development (2023), *Working age population (indicator).* doi:10.1787/d339918b-en. https://data.oecd.org/pop/working-age-population.htm

[39] WBOD, World Bank Open Data. (2023). global development data. Last accessed 06/01/2023. https://data.worldbank.org/

[40] Feenstra, R. C., Inklaar, R. & Timmer, M.P. (2022), The Next Generation of the Penn World Table. *American*

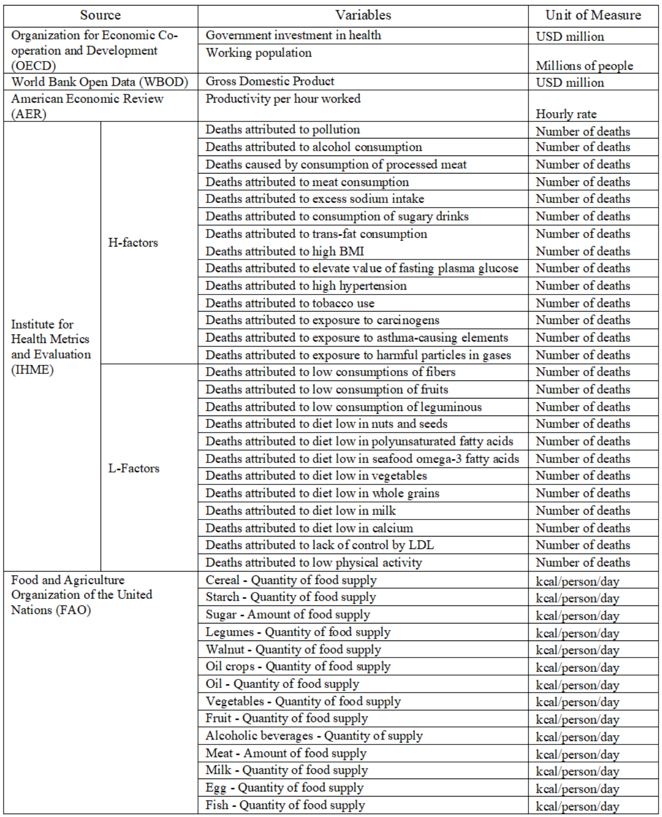
Economic Review, 105(10), 3150-3182. Last accessed 06/15/2023. https://www.rug.nl/ggdc/productivity/pwt/

[41] FAO, Food and Agriculture Organization of the United Nations. (2023). Food Balances. Last accessed 06/01/2023. https://www.fao.org/faostat/en/#data/FBSH

[42] IHME - Institute for Health Metrics and Evaluation (2020). *Global Burden of Disease Study 2019,* Results. https://vizhub.healthdata.org/gbd-results/.

[43] Shlens, J. (2014). A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*. https://doi.org/10.48550/arXiv.1404.1100

Annex 1. Databases and variables



Source		Variables	Unit of Measure
Organization for Economic Co-operation and Development (OECD)		Government investment in health	USD million
		Working population	Millions of people
World Bank Open Data (WBOD)		Gross Domestic Product	USD million
American Economic Review (AER)		Productivity per hour worked	Hourly rate
Institute for Health Metrics and Evaluation (IHME)	H-factors	Deaths attributed to pollution	Number of deaths
		Deaths attributed to alcohol consumption	Number of deaths
		Deaths caused by consumption of processed meat	Number of deaths
		Deaths attributed to meat consumption	Number of deaths
		Deaths attributed to excess sodium intake	Number of deaths
		Deaths attributed to consumption of sugary drinks	Number of deaths
		Deaths attributed to trans-fat consumption	Number of deaths
		Deaths attributed to high BMI	Number of deaths
		Deaths attributed to elevate value of fasting plasma glucose	Number of deaths
		Deaths attributed to high hypertension	Number of deaths
		Deaths attributed to tobacco use	Number of deaths
		Deaths attributed to exposure to carcinogens	Number of deaths
		Deaths attributed to exposure to asthma-causing elements	Number of deaths
		Deaths attributed to exposure to harmful particles in gases	Number of deaths
	L-Factors	Deaths attributed to low consumptions of fibers	Number of deaths
		Deaths attributed to low consumption of fruits	Number of deaths
		Deaths attributed to low consumption of leguminous	Number of deaths
		Deaths attributed to diet low in nuts and seeds	Number of deaths
		Deaths attributed to diet low in polyunsaturated fatty acids	Number of deaths
		Deaths attributed to diet low in seafood omega-3 fatty acids	Number of deaths
		Deaths attributed to diet low in vegetables	Number of deaths
		Deaths attributed to diet low in whole grains	Number of deaths
		Deaths attributed to diet low in milk	Number of deaths
		Deaths attributed to diet low in calcium	Number of deaths
		Deaths attributed to lack of control by LDL	Number of deaths
		Deaths attributed to low physical activity	Number of deaths
Food and Agriculture Organization of the United Nations (FAO)		Cereal - Quantity of food supply	kcal/person/day
		Starch - Quantity of food supply	kcal/person/day
		Sugar - Amount of food supply	kcal/person/day
		Legumes - Quantity of food supply	kcal/person/day
		Walnut - Quantity of food supply	kcal/person/day
		Oil crops - Quantity of food supply	kcal/person/day
		Oil - Quantity of food supply	kcal/person/day
		Vegetables - Quantity of food supply	kcal/person/day
		Fruit - Quantity of food supply	kcal/person/day
		Alcoholic beverages - Quantity of supply	kcal/person/day
		Meat - Amount of food supply	kcal/person/day
		Milk - Quantity of food supply	kcal/person/day
		Egg - Quantity of food supply	kcal/person/day
		Fish - Quantity of food supply	kcal/person/day

Source: OECD [28], WBOD [29], American Economic Review [30], IHME [31]; FAO [32]