

Analysis of life expectancy across countries using a decision tree

Ilknur Karacan,¹ Bahar Sennaroglu² and Ozalp Vayvay³

¹Department of Industrial Engineering, Institute of Pure and Applied Sciences, Marmara University, Istanbul, Turkey. ²Department of Industrial Engineering, Faculty of Engineering, Marmara University, Istanbul, Turkey (Correspondence to: Bahar Sennaroglu: bsennar@marmara.edu.tr). ³Department of Engineering Management, Institute of Pure and Applied Sciences, Marmara University, Istanbul, Turkey.

Abstract

Background: It is important to identify variables that influence life expectancy in order to develop strategies to improve health care systems and thereby increase life expectancy.

Aims: In this study, a decision tree was built using a chi-square automatic interaction detector technique in order to identify variables influencing life expectancy at birth.

Methods: Data were taken from the databases of the World Bank, World Health Organization and World Life Expectancy. Data from 166 countries for the year 2013 were extracted for 25 selected input variables related to mortality, health and the environment, child health, economy and demography in order to build the decision tree.

Results: Of the 25 variables, nine had a significant influence on life expectancy: percentage of the population using improved sanitation facilities; death rates per 100 000 population for HIV/AIDS, liver disease, stroke and coronary heart disease; percentage of the urban population using improved drinking-water sources; total fertility rate (births per woman); public health expenditure (percent of government expenditure); and health expenditure per capita.

Conclusions: Improving these variables may result in significant increases in life expectancy and quality of life. At the country level, appropriate strategies can be developed to improve the quality and performance of health care systems.

Keywords: life expectancy, decision trees, public health

Citation: Karacan I; Sennaroglu B; Vayvay O. Analysis of life expectancy across countries using decision tree. *East Mediterr Health J.* 2020;26(2):143–151. <https://doi.org/10.26719/2020.26.2.143>

Received: 24/08/16; accepted: 04/06/18

Copyright © World Health Organization (WHO) 2020. Open Access. Some rights reserved. This work is available under the CC BY-NC-SA 3.0 IGO license (<https://creativecommons.org/licenses/by-nc-sa/3.0/igo>).

Introduction

Life expectancy is the probable remaining life time of an individual after a specific age. Life expectancy at birth is the number of years, on average, a newborn can expect to live if the existing mortality rates continue to apply (1).

It is important to identify variables that influence and predict the life expectancy of the population of a country in order to develop appropriate strategies to improve the quality and performance of health care systems and thereby increase life expectancy.

Research to estimate life expectancy focuses on modelling life expectancy using time series or cross-sectional data based on trends in health and mortality observed in the population, as well as social, economic and environmental factors of the countries. Many studies have been conducted to estimate life expectancy going back to 1662 with an analysis to create a warning system for the onset, spread and decline of bubonic plague in London (2–13). Various modelling techniques have been used including an autoregressive integrated moving average (ARIMA) model (5), a novel model based on log mortality rate changes rather than levels (2), Bayesian spatiotemporal models (6), a model with a new survival function (7), Pollard's actuarial method of decomposing life expectancy (12) and extreme bounds analysis (13).

The determinants of life expectancy in Turkey (social, economic and environmental factors) were investigated

using time series data (4). The study concluded that while the availability of food and nutrition and health expenditure were the main determinant factors, smoking was the most important cause of death.

In this study, we aimed to build a decision tree using a chi-square automatic interaction detector (CHAID) technique to identify influential variables on life expectancy at birth.

Methods

Data sources

The data sources of this study are the databases of the World Bank, World Health Organization (WHO) and World Life Expectancy (14–16). World Life Expectancy is a free web portal developed by a private media company (16). The World Bank data meet Open Data best practices and the established standards of professional data communities (17). WHO provides access to health-related statistics for its 194 Member States with its Global Health Observatory data repository (18). Although the datasets are not always the same as official national estimates, they represent the best estimates of WHO using methodologies for specific indicators that aim for comparability across countries and time (18). Using these databases, data from 166 countries for the year 2013 were extracted for 25 selected input variables in order to build the decision tree to explain life expectancy at birth. The variables

were grouped into five categories related to mortality, health and environment, child health, economy and demography (Box 1).

Decision tree

The decision tree technique is an intuitive data mining tool that is capable of handling heterogeneous data by defining explicit rules for classification (19). The technique splits the data to determine the most significant predictor recursively that yields the best separation at the current level. The aim is to divide data into homogenous sublevels. The decision tree is transparent and has easy to understand solutions provided by the algorithm which helps focus on the entire data (20).

A tree diagram is a useful tool to visualize the structure of a decision tree. Each path from the root node to one of the leaf nodes in the decision tree provides a rule for classification. Therefore, it is easy to interpret the rules from a decision tree. Other advantages of decision trees are: (i) they do not require any particular probability distributions for variables and work with both discrete and continuous variables, (ii) they are not affected by collinearity and (iii) they are insensitive to outliers in the data set. However, decision trees suffer from problems of statistical reliability and generalizability. They can also bias classification choices because of sequential variable evaluation. A reduced data set is evaluated after each split in the decision tree which introduces bias.

Chi-square automatic interaction detector

We built a decision tree using a CHAID technique. The target variable was life expectancy at birth (LIFEX), and the 25 selected variables were used as the input variables of the decision tree.

CHAID is a popular decision tree technique that was described in 1975 and formalized in 1980 (19). The

algorithm works on nodes to build a hierarchical tree. It decides the most significant predictor to split the data iteratively using the chi-square test. CHAID contains several components on a decision tree. The root node, at the top of the hierarchy, involves the dependent or target variable. In our study, life expectancy at birth stayed at the root as the target variable to be explained by the input variables. Parent nodes are the nodes that are composed by dividing the root node by the most significant predictor. In our study, access to sanitation facilities appeared to be the most significant predictor that divided the data into five nodes. The child nodes are the successor nodes for parent nodes on the decision tree. The terminal nodes, also called leaf nodes, are the final nodes on the decision tree.

The steps of the CHAID algorithm are explained in Data mining and statistics for decision making (19). CHAID makes use of the chi-square test by merging classes that do not have significantly different effects on the target variable, and then chooses the best split and decides whether it is worth performing any additional splits on a node (21). Four steps of CHAID are for merging the categories and the last step is for node splitting (19).

1) For each input variable with more than two categories, the chi-square test is done to group the categories by cross-tabulating them with the categories of the target variable. In subcross-tabulations, the pair of categories of the input variable with the smallest chi-square value (the largest *P*-value) is selected and compared with the chosen threshold (default value is $\alpha = 0.05$). If the chi-square is not significant (*P*-value greater than the chosen threshold), the two categories are merged.

2) Step 1 is repeated until all the pairs of categories have a significant chi-square (*P*-value less than the chosen threshold) or until there are only two categories remaining for each input variable.

Box 1 Variables selected to build the decision tree

Mortality-related variables:
Death rates per 100 000 population for all cancers (CANCER), coronary heart disease (CHD), stroke (STROKE), HIV/AIDS (HIV_AIDS), lung disease (LUNGD), kidney disease (KD), hypertension (HTEN), liver disease (LIVERD), alcohol use (ALCOHOL), drug use (DRUG), and road traffic accidents (RTA)
Health and environment-related variables:
Percentage of the urban population using improved drinking-water sources (DWS_U), percentage of the rural population using improved drinking-water sources (DWS_R), percentage of the population using improved sanitation facilities (SF), carbon dioxide emissions as metric tons per capita (CO ₂), forest area as percentage of land (FOREST), and food safety as a percent of the average of all indicators that reflect the level of performance or achievement of the International Health Regulations Food Safety (FOOD)
Child health-related variables:
Measles immunization (ImMeasles), and diphtheria, pertussis and tetanus immunization (ImDPT) as the percent of children immunized aged 12–23 months
Economy-related variables:
Total health expenditure as a percent of gross domestic product (HEoGDP), public health expenditure as a percent of total health expenditure (HEoTHE), public health expenditure as a percent of government expenditure (HEoGE), and health expenditure in US\$ per capita (HEpC)
Demography-related variables:
Total fertility rate as births per woman (FR) and population growth as annual percent (PG)

3) If the input variable is nominal and has missing values, the set of missing values is considered to be a category and treated in the same way as the others. If the input variable is ordinal or quantitative, the missing values category is merged with another category with the closest chi-square after the end of the preceding merger processes.

4) The *P*-value associated with the chi-square of the best table obtained when the merging process stops is multiplied by the Bonferroni correction (adjusted *P*-value) in order to prevent the over-evaluation of the significance of the multiple-category variables.

5) When the categories have been grouped optimally for each input variable and the adjusted *P*-value has been calculated, CHAID selects the variable for which the chi-square is most significant (the one for which the adjusted *P*-value is smallest). If the adjusted *P*-value is less than the chosen threshold for the split (default value is $\alpha = 0.05$), the node is divided into a number of child nodes equal to the number of categories of the variable after grouping, otherwise the node is not divided.

Results

Using the CHAID decision tree technique, of all the 25 input variables analysed, nine were identified as the most significant variables affecting life expectancy at birth. These variables were: percentage of the population using improved sanitation facilities; death rates per 100 000 population for HIV/AIDS, liver disease, stroke and coronary heart disease; percentage of the urban population using improved drinking-water sources; total fertility rate as births per woman; public health expenditure as a percent of government expenditure; and health expenditure as US\$ per capita.

As seen in Figure 1, the percentage of the population using improved sanitation facilities was the best first split. The percentage of population using improved sanitation facilities divides the entire group of countries into five groups in terms of life expectancy (Table 1). If the percentage of the population using improved sanitation facilities is at most 59%, this group includes 51 countries and their average life expectancy is 60.564 years (Node 1), whereas if the percentage of the population using improved sanitation facilities is more than 99%, this group includes 20 countries and their average life expectancy is 80.664 years (Node 5). Based on the World Bank's classification of countries by region and income level, Node 1 contains 27 low-income countries and 24 middle-income countries from sub-Saharan Africa, South Asia, Middle East and North Africa, Latin America and Caribbean, and East Asia and Pacific regions. Node 5 contains 19 high-income countries and one middle-income country from Middle East and North Africa, East Asia and Pacific, Europe and Central Asia, and North America regions. Among the Member States of the WHO Eastern Mediterranean Region, four countries are in Node 1 (Sudan, Afghanistan, Djibouti and Yemen), one country is in Node 2 (Pakistan), seven

are in Node 3 (Egypt, Islamic Republic of Iran, Iraq, Lebanon, Morocco, Syria and Tunisia), six are in Node 4 (Jordan, Libya, Bahrain, Oman, Qatar and United Arab Emirates) and two countries are in Node 5 (Kuwait and Saudi Arabia).

The CHAID procedure takes each of these five groups and finds the variable that best splits each group. From the root node to the leaf nodes, 166 countries were divided into 26 mutually exclusive groups. The group with the lowest life expectancy (55.667 years) included countries in sub-Saharan Africa. These countries are Angola, Burundi, Central African Republic, Chad, Democratic Republic of the Congo, Guinea, Guinea-Bissau, Liberia, Malawi, Sierra Leone, Togo, Uganda, which are on the list of least developed countries and Cameroon, the Congo, Côte d'Ivoire, Ghana, Nigeria, Kingdom of Eswatini and Zambia which are middle-income countries. For this group, the decision tree gave the following rule: $SF \leq 59\%$ of the population and $HIV_AIDS > 48.87$ deaths per 100 000 population and $LIVERD > 24.34$ deaths per 100 000 population then LIFEX is predicted as 55.667 years.

The group with the highest life expectancy of 82.442 years included Australia, Canada, Cyprus, Israel, Italy, Japan, Malta, New Zealand, Singapore, Spain and Switzerland which are all high-income countries. For this group the decision tree gave the following rule: $SF > 99\%$ of the population and $DWS_U > 99\%$ and $LIVERD \leq 7.1$ deaths per 100 000 population then LIFEX is predicted as 82.442 years.

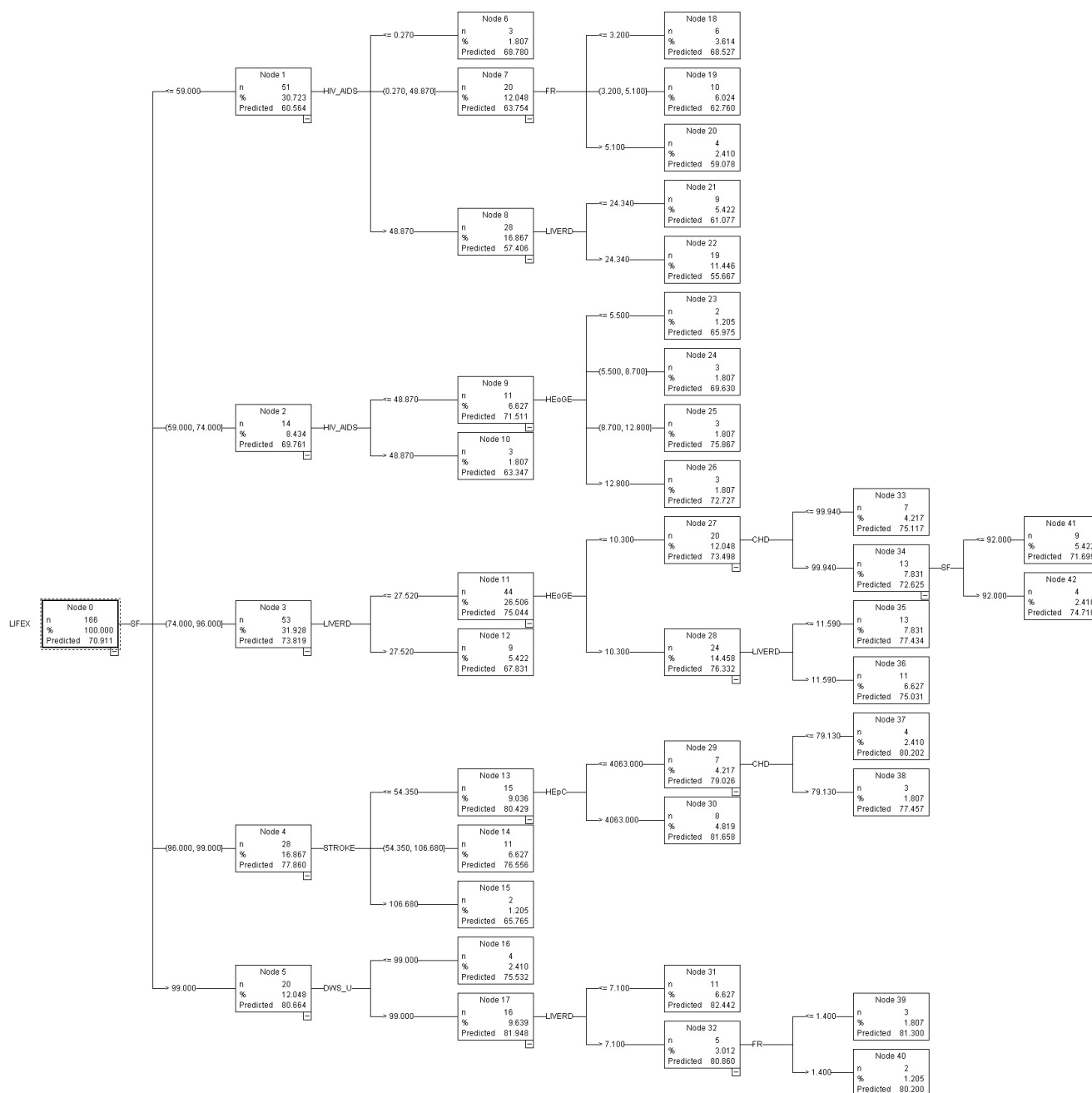
The groups with life expectancy less than 60 years include countries in the sub-Saharan Africa region which are mostly least developed countries. If life expectancy is about 70 years, groups contain middle-income countries. If life expectancy is more than 81 years, high-income countries are included which are Austria, Portugal and the Republic of Korea. For this group, the decision rule is: $SF > 99\%$ of population and $DWS_U > 99\%$ and $LIVERD > 7.1$ deaths per 100 000 population and $FR \leq 1.4$ births per woman then LIFEX is predicted as 81.3 years.

For the group including high-income countries Belgium and Denmark, fertility rate also has an effect and the decision rule is: $SF > 99\%$ and $DWS_U > 99\%$ and $LIVERD > 7.1$ deaths per 100 000 population and $FR > 1.4$ births per woman then LIFEX is predicted as 80.2 years

This life expectancy is lower than the previous group. Higher fertility rates cause an increase in the population and since the available resources and infrastructure will therefore be shared by more individuals, the quality of life and life expectancy will be adversely affected.

Table 2 shows the groupings of Member States of the WHO Eastern Mediterranean Region for the three lowest and highest life expectancy predictions. Decision rules indicate that an increase in the percentage of the population using improved sanitation facilities will increase life expectancy. Of the mortality-related variables, the death rates for HIV/AIDS, liver disease, stroke and coronary heart disease significantly influence life expectancy for these countries. Economy-related

Figure 1 Decision tree for life expectancy at birth



LIFEX: life expectancy at birth; SF: percentage of the population using improved sanitation facilities; HIV_AIDS: death rate per 100 000 population for HIV/AIDS; LIVERD: death rate per 100 000 population for liver disease; STROKE: death rate per 100 000 population for STROKE; DWS_U: percentage of the urban population using improved drinking-water sources; FR: total fertility rate as births per woman; HeoGE: public health expenditure as a percent of government expenditure; HepC: health expenditure in US\$ per capita; CHD: death rate per 100 000 population for coronary heart disease.

variables are also important and it is clear that more government spending on public health will result in increased life expectancy. Since decision rules for high-income countries with high life expectancy indicate that the percentage of the urban population using improved drinking-water sources is very high for these countries, this variable should also be taken into consideration by countries of the Eastern Mediterranean Region in order to increase life expectancy.

Discussion

The results of this study are consistent with those of previous studies. Research using extreme bound analysis identified eight variables as robust predictors for life expectancy: improved water sources, poverty headcount rate, adolescent fertility rate, labour participation rate, health aid, share of pregnant women receiving prenatal care, and the Country Policy and Institutional Assessment gender equality index (10). Other research ascertained

Table 1 Groups of countries according to percentage of the population using improved sanitation facilities (SF)

		Node 1	Node 2	Node 3	Node 4	Node 5
		SF ≤ 59%	59% < SF ≤ 74%	74% < SF ≤ 96%	96% < SF ≤ 99%	SF > 99%
Number of countries		51	14	53	28	20
Life expectancy (years)		60.564	69.761	73.819	77.86	80.664
Region	Income level	Countries in the node ^a				
Sub-Saharan Africa	Low income	Benin, Burkina Faso, Burundi, Central African Republic, Chad, Comoros, Democratic Republic of the Congo, Eritrea, Ethiopia, Gambia, Guinea, Guinea-Bissau, Liberia, Madagascar, Malawi, Mali, Mozambique, Namibia, Niger, Senegal, Sierra Leone, United Republic of Tanzania, Togo, Uganda	Rwanda			
	Middle income	Angola, Cameroon, Republic of Congo, Côte d'Ivoire, Gabon, Ghana, Kenya, Lesotho, Mauritania, Nigeria, Sudan , Kingdom of Eswatini, Zambia	Botswana, Cabo Verde, South Africa	Equatorial Guinea, Mauritius		
South Asia	Low income	Afghanistan , Nepal				
	Middle income	Bangladesh, Bhutan, India	Pakistan	Sri Lanka	Maldives	
Middle East and North Africa	Middle income	Djibouti, Yemen		Algeria, Egypt, Iran (Islamic Republic of), Iraq, Lebanon, Morocco, Syrian Arab Republic, Tunisia	Jordan, Libya	
	High income				Bahrain, Oman, Qatar, United Arab Emirates	Israel, Kuwait , Malta, Saudi Arabia
Latin America and Caribbean	Low income	Haiti				
	Middle income	Bolivia (Plurinational State of)	El Salvador, Guatemala, Nicaragua, Panama	Argentina, Belize, Brazil, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, Guyana, Honduras, Jamaica, Mexico, Paraguay, Peru, Suriname, Venezuela (Bolivarian Republic of)		
	High income			Bahamas, Barbados, Trinidad and Tobago, Uruguay	Chile	

Table 1 Groups of countries according to percentage of the population using improved sanitation facilities (SF) (concluded)

Region	Income level	Countries in the node ^a				
East Asia and Pacific	Middle income	Cambodia, Mongolia, Papua New Guinea, Solomon Islands, Timor-Leste	China, Indonesia, Lao People's Democratic Republic, Philippines	Fiji, Malaysia, Myanmar, Thailand, Viet Nam		
	High income					Australia, Japan, New Zealand, Singapore, Republic of Korea
Europe and Central Asia	Middle income		Russian Federation	Albania, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Georgia, Kyrgyzstan, North Macedonia, Moldova, Romania, Tajikistan, Turkey, Ukraine	Croatia, Kazakhstan, Serbia, Turkmenistan	Uzbekistan
	High income			Ireland, Latvia, Lithuania	Czechia, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Luxembourg, Netherlands, Norway, Poland, Slovakia, Slovenia, Sweden, the United Kingdom	Austria, Belgium, Cyprus, Denmark, Italy, Portugal, Spain, Switzerland
North America	High income					Canada, United States of America

^aMember States of the World Health Organization Eastern Mediterranean Region are in bold text.

that falling tobacco use for men and a decrease in cardiovascular disease mortality for both men and women were the main factors contributing to increase in life expectancy in older age in high-income countries (11). The findings indicated that progress in the improvement rate in older age mortality has been slower in low- and middle-income countries because of continuing communicable disease epidemics, such as HIV/AIDS and tuberculosis, and the growing epidemic of noncommunicable diseases. A recent study concluded that decrease in deaths from cardiovascular disease was the most important contributor to the change in life expectancy (12), while another identified the most influential determinants of longer life expectancy at birth in low-income countries were HIV prevalence among children, gender equality, agricultural production, political stability, improved water source, improved sanitation facilities, good governance, primary school enrolment, increased private health expenditure and overseas development assistance, and control of armed conflict and HIV prevalence among men (13).

In our study, the percentage of the population using improved sanitation facilities was the most significant variable affecting life expectancy. The difference between the groups of countries in the decision tree with the highest life expectancy (node 5) and the lowest life expectancy (node 1) was about 20 years, which indicates that improved sanitation facilities increases life

expectancy. A comprehensive study on sanitation and health supports this finding (22); it stated that 2.6 billion people in the world do not have adequate sanitation and inadequate sanitation causes about 10% of global diseases.

HIV/AIDS was the next significant variable in grouping the countries with a low percentage of the population having improved sanitation facilities (nodes 1 and 2). Most of the countries in these groups (63%) (nodes 6, 7 and 8 under node 1, and nodes 9 and 10 under node 2; Figure 1) are low- and middle-income sub-Saharan countries (Figure 1 and Table 1). Sub-Saharan Africa accounts for 71% of the global burden of HIV infection, although it is home to only 12% of the global population (23). Policies directed at preventive actions and treatment programmes will help reduce deaths from HIV/AIDS and increase life expectancy. Similarly, policies directed at reducing deaths from liver disease, stroke and coronary heart disease, which were also identified as influential variables, will have positive effects on life expectancy. The decision rules generated by the CHAID decision tree technique provide easy interpretation of differences between groups of countries and hence can be used in strategy development.

A limitation of this study is that only 25 variables related to mortality, health and environment, child health, economy, and demography, were included whereas other

Table 2 Member States of the World Health Organization Eastern Mediterranean Region with lowest and highest life expectancy

Member State at the node	Other countries at the node	Node	Decision rule for the node	Predicted life expectancy for the node (years)
Djibouti	Ethiopia, Gabon, Haiti, Kenya, Lesotho, Mozambique, Namibia, United Republic of Tanzania	21	SF ≤ 59% of population and HIV ₋ AIDS > 48.87 deaths per 100 000 population and LIVERD ≤ 24.34 deaths per 100 000 population	61.077
Afghanistan, Sudan, Yemen	Benin, Comoros, Eritrea, Madagascar, Mauritania, Papua New Guinea, Senegal	19	SF ≤ 59% of population and 0.27 < HIV ₋ AIDS ≤ 48.87 deaths per 100 000 population and 3.2 < FR ≤ 5.1 births per woman	62.760
Pakistan	Lao People's Democratic Republic	23	59 < SF ≤ 74% of population and HIV ₋ AIDS ≤ 48.87 deaths per 100 000 population and HeoGE ≤ 5.5% of government expenditure	65.975
Iran (Islamic Republic of), Lebanon, Tunisia	Bahamas, Barbados, Colombia, Costa Rica, Cuba, Ireland, North Macedonia, Suriname, Turkey, Uruguay	35	74 < SF ≤ 96% of population and LIVERD ≤ 27.52 deaths per 100 000 population and HEoGE > 10.3% of government expenditure and LIVERD ≤ 11.59 deaths per 100 000 population	77.434
Bahrain	Czechia, Estonia	38	96 < SF ≤ 99% of population and STROKE ≤ 54.35 deaths per 100 000 population and HEpC ≤ US\$ 4063 per capita and CHD > 79.13 deaths per 100 000 population	77.457
Qatar	Chile, Slovenia, United Kingdom	37	96 < SF ≤ 99% of population and STROKE ≤ 54.35 deaths per 100 000 population and HEpC ≤ US\$ 4063 dollars per capita and CHD ≤ 79.13 deaths per 100 000 population	80.202

SF: percentage of the population using improved sanitation facilities; HIV₋AIDS: death rate per 100 000 population for HIV/AIDS; LIVERD: death rate per 100 000 population for liver disease; FR: total fertility rate as births per woman; HeoGE: public health expenditure as a percent of government expenditure; STROKE: death rate per 100 000 population for stroke; HepC: health expenditure in US\$ per capita; CHD: death rate per 100 000 population for coronary heart disease.

variables that were not considered may also influence life expectancy at birth, such as level of education and employment by education level.

In addition, we used cross-sectional data for 2013 without gender segregation. Taking account of different years and gender in the analysis requires the construction of separate decision trees which might result in different decision rules for classification. However, since the results of our study are consistent with those of the similar previous studies, decision trees are proposed as an effective method for the analysis of life expectancy and for health research in general. The most important variables identified by our study can also be used as input variables in other similar studies.

Conclusion

Based on our CHAID decisions tree analysis, life expectancy at birth varied greatly between countries. Our results indicate that the factors significantly affecting life expectancy at birth are: (i) mortality-related variables – death rates per 100 000 population for HIV/AIDS, liver disease, stroke and coronary heart disease; (ii) health and environment-related variables – percentage of the

population using improved sanitation facilities and percentage of the urban population using improved drinking-water sources; (iii) economy-related variables – public health expenditure as a percent of government expenditure and health expenditure in US\$ per capita; and (iv) demography-related variables – total fertility rate as births per woman. Because high-income countries have improved these factors substantially, their populations have longer life expectancy than people in low- and middle-income countries.

Targeting the variables found to influence life expectancy may result in significant increases in life expectancy and quality of life. At the country level, appropriate strategies can be developed to improve the quality and performance of health care systems. Efforts to improve life expectancy may require budget allocation and/or appropriate regulations.

For future work, new input variables can be added to the analysis and another decision tree technique, such as classification and regression tree, can be applied.

Funding: None.

Competing interests: None declared.

Analyse de l'espérance de vie dans différents pays au moyen d'un arbre de décision

Résumé

Contexte : Il est important d'identifier les variables influençant l'espérance de vie pour mettre au point des stratégies d'amélioration des systèmes de santé et faire ainsi augmenter l'espérance de vie.

Objectifs : Dans la présente étude, un arbre de décision a été établi à l'aide d'une technique automatique de détection des interactions selon le test du khi-carré, afin d'identifier les variables influençant l'espérance de vie à la naissance.

Méthodes : Les données provenaient des bases de données de la Banque mondiale, de l'Organisation mondiale de la Santé et du portail Web « World Life Expectancy ». Les données de 166 pays pour l'année 2013 ont été extraites par rapport à une sélection de 25 variables d'entrée liées à la mortalité, la santé et l'environnement, la santé de l'enfant, l'économie et la démographie, afin d'établir l'arbre de décision.

Résultats : Neuf des 25 variables avaient une influence importante sur l'espérance de vie : le pourcentage de la population ayant accès à des installations sanitaires améliorées ; le taux de mortalité pour 100 000 habitants lié au VIH/sida, aux maladies hépatiques, aux accidents vasculaires cérébraux et aux coronaropathies ; le pourcentage de la population urbaine ayant accès à des sources d'eau de boisson améliorées ; le taux de fécondité total (naissances par femme) ; les dépenses publiques de santé (pourcentage des dépenses publiques) ; et les dépenses de santé par habitant.

Conclusions : L'amélioration de ces variables pourrait se traduire par des augmentations significatives de l'espérance de vie et de la qualité de vie. À l'échelle des pays, des stratégies appropriées peuvent être mises au point pour améliorer la qualité et le fonctionnement des systèmes de santé.

تحليل مأمول العمر على مستوى البلدان باستخدام شجرة القرارات

كنور كارجان، باهار سنار أوجلو، أوزالب فاي فايفاي

الخلاصة

الخلفية: من المهم تحديد المتغيرات التي تؤثر على مأمول العمر حتى يتسنى إعداد الاستراتيجيات اللازمة لتحسين نظم الرعاية الصحية، ومن ثم، زيادة معدل مأمول العمر.

الأهداف: في هذه الدراسة، بُنيت شجرة قرارات باستخدام طريقة² لاكاشف التفاعل التلقائي لتحديد المتغيرات المؤثرة على مأمول العمر عند الميلاد.

طرق البحث: أُخذت بيانات من قواعد بيانات البنك الدولي ومنظمة الصحة العالمية وموقع «World Life Expectancy» كذلك استُخرجت بيانات من 166 بلداً لعام 2013 بشأن 25 متغيراً من متغيرات المدخلات المختارة والمربطة بالوفاة، والصحة، والبيئة، وصحة الطفل، والاقتصاد، والسكانية، وذلك لاستخدامها في بناء شجرة القرارات.

النتائج: من بين المتغيرات الخمسة والعشرين، كان لتسعة منها تأثير كبير على مأمول العمر: نسبة السكان التي تستخدم مرافق إصحاح مُحسّنة؛ ومعدلات الوفيات لكل 100000 نسمة بسبب فيروس العوز المناعي البشري/الإيدز، وأمراض الكبد، والسكتات الدماغية، وأمراض القلب التاجية؛ ونسبة سكان المناطق الحضرية الذين يستخدمون مصادر مُحسّنة لمياه الشرب؛ ومعدل الخصوبة الكلي (عدد الولادات لكل سيدة)؛ ونفقات الصحة العامة (نسبة النفقات الحكومية)، والنفقات الصحية للفرد.

الاستنتاجات: قد يؤدي تحسين المتغيرات إلى تحقيق زيادة مأمول العمر وتحسين جودة الحياة بصورة كبيرة. وعلى المستوى القطري، يمكن إعداد استراتيجيات ملائمة لتحسين جودة الرعاية الصحية ومستوى أداؤها.

References

1. WHO statistical information system [online database]. Life expectancy at birth. Geneva: World Health Organization; 2016 (<https://www.who.int/whosis/whostat2006DefinitionsAndMetadata.pdf>, accessed 4 January 2016).
2. Mitchell D, Brockett P, Mendoza-Arriaga R, Muthuraman K. Modeling and forecasting mortality rates. *Insur Math Econ.* 2013;52(2):275–85.
3. Kinsella K, Velkoff V. Life expectancy and changing mortality. *Aging Clin Exp Res.* 2002;5(14):322–32.
4. Halicioglu F. Modeling life expectancy in Turkey. *Econ Model.* 2011;28(5):2075–82.
5. Torri T, Vaupel JW. Forecasting life expectancy in an international context. *Int J Forecas.* 2012;28(2):519–31.
6. Bennett JE, Li G, Foreman K, Best N, Kontis V, Pearson C, et al. The future of life expectancy and life expectancy inequalities in England and Wales: Bayesian spatiotemporal forecasting. *Lancet.* 2015;386(9989):163–70.
7. Wong CH, Tsui AK. Forecasting life expectancy: evidence from a new survival function. *Insur Math Econ.* 2015;65:208–26.

8. Lee RD, Carter LR. Modeling and forecasting US mortality. *J Am Stat Assoc.* 1992;87(419):659–71.
9. Bongaarts J. Long-range trends in adult mortality: models and projection methods. *Demography.* 2005;42(1):23–49.
10. Carmignani F, Shankar S, Tan EJ, Tang KK. Identifying covariates of population health using extreme bound analysis. *Eur J Health Econ.* 2014;15(5):515–31.
11. Mathers CD, Stevens GA, Boerma T, White RA, Tobias MI. Causes of international increases in older age life expectancy. *Lancet.* 2015;385(9967):540–8.
12. Klenk J, Keil U, Jaensch A, Christiansen MC, Nagel G. Changes in life expectancy 1950–2010: contributions from age- and disease-specific mortality in selected countries. *Popul Health Metr.* 2016;14(1):20–30.
13. Hauck K, Martin S, Smith PC. Priorities for action on the social determinants of health: Empirical evidence on the strongest associations with life expectancy in 54 low-income countries, 1990–2012. *Soc Sci Med.* 2016;167:88–98.
14. World Bank. Data [online database] Washington (DC): World Bank; 2013 (<https://databank.worldbank.org/source/world-development-indicators>, accessed 27 June 2019).
15. Global Health Observatory data repository. World Health Statistics [online database]. Geneva: World Health Organization; 2013 (<https://apps.who.int/gho/data/node.main.1?lang=en>, accessed 4 January 2016).
16. World Life Expectancy Map [online database]. World Life Expectancy; 2013 (<http://www.worldlifeexpectancy.com/world-life-expectancy-map>, accessed 4 January 2016).
17. World Bank. Data. Supply and quality of data. Washington (DC): World Bank; 2018 (<http://opendatatoolkit.worldbank.org/en/supply.html>, accessed 26 March 2018).
18. Global health observatory resources. About the observatory. Geneva: World Health Organization; 2018 (<http://apps.who.int/gho/data/node.resources>, accessed 26 March 2018).
19. Tufféry S. Data mining and statistics for decision making. Chichester: John Wiley & Sons Ltd; 2011: 325–7.
20. Tsiptsis KK, Chorianopoulos A. Data mining techniques in CRM: inside customer segmentation. Chichester: John Wiley & Sons Ltd; 2009.
21. Berry MJA, Linoff GS. Data mining techniques: for marketing, sales, and customer relationship management, second edition. Indianapolis, Wiley Publishing, Inc.; 2004.
22. Mara D, Lane J, Scott B, Trouba D. Sanitation and health. *PLoS Med.* 2010;7(11):e1000363.
23. Kharsany AB, Karim QA. HIV infection and AIDS in Sub-Saharan Africa: current status, challenges and opportunities. *Open AIDS J.* 2016;10:34–48.