Available online at www.alphanumericjournal.com

alphanumeric journal

The Journal of Operations Research, Statistics, Econometrics and Management Information Systems

Volume 9, Issue 2, 2021

Received: May 21, 2021
Accepted: July 13, 2021
Published Online: December 31, 2021

AJ ID: 2021.09.02.OR.03
DOI: 10.17093/alphanumeric.940744
Research Article

The Determinants of Turkey Household Catastrophic Health Expenditures: A Different Approach by Data Mining

Uğur Ercan, Ph.D.

Lect., Department of Informatics, Akdeniz University, Antalya, Turkey, ugurercan@akdeniz.edu.tr

* Akdeniz Üniversitesi Enformatik Bölüm Başkanlığı, Akdeniz Üniversitesi, İletişim Fakültesi Yeni Binası, Kampüs, Antalya, Türkiye

ABSTRACT

This study aims to determine socioeconomic, demographic, and household characteristics that affect Turkish household catastrophic health expenditure (CHE). Data gathered by TurkStat belonging to 40,033 households for the years 2009-2012 were used in the analysis. In the study, CHE was defined as household health expenditures that were 40% (or above) greater than the capacity pay of the household. CHAID analysis was used to determine characteristics affecting Turkish household CHE. According to the CHAID analysis; income, presence of a sick/disabled person, residential area, household size, age, education level and gender of the household head, presence of individuals aged 65+, presence of people between the ages of zero and five, and access to health institutions have been observed to affect CHE, while the marital status, age, and employment status of the head of the household, or the household type do not affect CHE. The proportion of households exposed to CHE was 0.62%, and the proportion of households making out-of-pocket health expenditure (OOPHE) was 62.71%. It was especially observed that households with low income, with sick/disabled individuals, and those with difficult access to healthcare facilities are more likely to be exposed to CHE. None of the households exposed to CHE has supplementary health insurance.

Keywords: Catastrophic Health Expenditure, Out-of-pocket Health Expenditure, Household budget survey, Turkey, CHAID Analysis

1. Introduction

The world is experiencing economic, environmental, technological and demographic changes that affect well-being and health. Of these, economic growth is directly related to improving health and well-being, that are both cause and effect (WHO & UNICEF, 2018). Most countries have experienced economic growth in the last 20 years, the economy growing by 3% between 2000 and 2017 (WHO, 2019). The sustainable development agenda for 2030 includes many important goals such as achieving universal health insurance, promoting physical and mental health and wellbeing, providing access to quality healthcare, extending life expectancy, access to safe, effective and quality basic pharmaceuticals and vaccines, and financial risk protection (UN, 2015). Despite these targets, health expenditures (HE) are gradually increasing. Between 2000 and 2017, global HE grew by 3.9% per year, while HE rose from $7.6 trillion in 2016 to $7.8 trillion in 2017 (WHO, 2019). HE has outpaced economic growth in the past. It is expected to do so in the future, despite the slowing in recent years (OECD, 2019).

Democratic Republic of Congo has the lowest HE per capita in 2017 at $ 19.434, the other countries respectively, Mozambique (≈$21), Ethiopia (≈$25), and Rwanda (≈$49). Kiribati has the lowest OOPHE per capita in 2017 at $0.19, the other countries respectively, Mozambique (≈$1.56), Rwanda (≈$3.07), Democratic Republic of Congo (≈$7.80), and Ethiopia (≈$8.69). Countries with the lowest spending in both HE per capita and OOPHE per capita, except Kiribati, are Sub Saharan African countries. The United States of America has the highest HE per capita in 2017 at $10246.14, the other countries respectively, Switzerland (≈$9956), Norway (≈$7936), Australia (≈$5332). Switzerland has the highest per capita OOPHE in 2017, at $2882.04, the other countries respectively, The USA (≈$1126), Norway (≈$1125), Australia (≈$968) (World Bank, 2021a; World Bank, 2021b).

In terms of income groups, it is seen that HE per capita, and OOPHE per capita is very low in all groups except the high-income group. According to World Bank data, Turkey was located in the upper-middle-income group, in 2017, with a HE per capita of about $445, while OOPHE per capita was about $78 (World Bank, 2021a; World Bank, 2021b).

Across the world, while approximately 150 million people face CHE every year, therebeside 100 million people are dragged below the poverty line (Xu et al., 2007). It has been stated in previous studies that CHE are frequently seen in countries with poverty and inequality (Xu et al., 2003; Falconi & Bernabé, 2018; Njagi et al., 2018). According to a study carried out in Caribbean and Latin American countries, the proportion of households exposed to CHE ranged between 1% and 25% (Knaul et al., 2011). In a study examining the studies conducted in sub-Saharan Africa, it was observed that CHE were high in sub-Saharan African countries, especially in West African countries (Njagi et al., 2018). CHE can also be seen in countries developed in health care services that have advanced technology, and high income per capita. For example; according to a study conducted in Korea, 2.1% to 2.5% of households faced CHE (Kang & Kim, 2018). The catastrophic impact is not only seen in low-income and middle-income countries but also in OECD countries (Arsenijevic et al., 2013). Among the OECD countries, less than 2% of households in France, Sweden, the United

Kingdom, Ireland, Czech Republic and Slovenia (OECD, 2019) face it. When viewed on studies conducted in Turkey, the percentage of households that face CHE is averagely 0.49% (between 2002-2014) in the study of Tokatlıoğlu & Tokatlıoğlu (2018), the study of Yardim et al. (2010) were seen as 0.60% in 2006, the sudy of Yereli et al. (2014) were seen as 0.30% in 2011, and the study of Narcı et al. (2015) were seen as 0.75% in 2010.

The most important effect of the 2008 world economic crisis on economies is the shrinkage of production. There is a serious relationship between economic indicators such as economic situation, unemployment, inflation and GNP and economic crisis. In addition to its economic effects, crisis also has social and political effects. Health sector is one of the areas affected by it. Since health and health services have a determining role in the economy, a mutual relationship between health and economy can be mentioned. Turkey has taken various measures to reduce the effects of the world economic crisis (Memişoğlu & Durgun, 2011). In the 2009-2011 period, it was planned to give more importance to education, health, and social expenditures and reduce regional development disparities. In this context, it was planned to increase the quality of life standards of the society, improve income distribution, expand preventive health services, facilitate access, and increase quality (T.C. Maliye Bakanlığı, 2010). After the 2008 world economic crisis, the study carried out using household data between 2009 and 2012 is important in this respect.

This study aims to determine socioeconomic, demographic, and household characteristics that affect Turkish household CHE using data mining techniques. Econometric models such as logistic regression (binomial, multinomial), quantile regression, poisson regression, Heckmann model, and zero-inflated negative binomial have been used in almost all of the studies in the literature. CHAID analysis, one of the data mining methods that have not been used in the literature before, was used to determine the CHE of households. This method used is the most important contribution of the conducted study to the literature. The study consists of five parts: introduction, literature, material and method, findings, and conclusion.

2. Literature

Health expenditures are defined as catastrophical if HE exceeds a portion (threshold value) of total expenditure or household income in a given period, usually within a year (O'Donnell et al., 2008). There is no consensus on the threshold value. It can vary from 5% to 40% of household income or expenditure (Wagstaff & van Doorslaer, 2003; O'Donnell et al., 2008; Goryakin & Suhrcke, 2014; Wagstaff et al., 2018). According to WHO (World Health Organization), if the household OOPHE is equal to or more than 40 percent of the household's payment capacity, these households are defined as those exposed to CHE (Xu, 2005; Xu et al., 2005). Household's capacity to pay, defined as the difference left by subtracting basic subsistence expenditure from total household expenditure (Xu et. al., 2003; Xu, 2005; Xu et al., 2007). In this study, the method accepted by WHO was used to determine whether households made CHE. How to calculate the subsistence expenditure, poverty line and capacity to pay, and how to determine the households exposed to CHE were stated by Xu and colleagues (Xu et. al., 2003; Xu, 2005; Xu et al., 2007).

Studies conducted to examine the determinants of CHE stated that many variables such as socioeconomic, sociodemographic, and individual characteristics could affect catastrophe. The countries where studies were conducted are given in the paranthesis (Yardim et al., 2010, (Turkey); Barros et al., 2011, (Brazil); Arsenijevic et al., 2013, (Serbia); Li et al., 2013, (China); Yereli et al., 2014, (Turkey); Choi et al., 2015, (Korea); Narcı et al., 2015, (Turkey); Rashad & Sharaf, 2015, (Egypt); Piroozi et al., 2016, (Iran); Ahmed et al., 2018, (Vietnam); Cleopatra & Eunice, 2018, (Nigeria); Falconi & Bernabé, 2018, (Peru); Tokatlıoğlu & Tokatlıoğlu, 2018, (Turkey); Si et al., 2019, (China); Akhtar et al., 2020, (India); Dalui et al., 2020, (India); Vahedi et al., 2020, (Iran); Zhao et al., 2020, (China); Thu Thuong et al., 2021, (Vietnam)).

When we look at these characteristics in detail; the variables listed below have been shown to have an impact;

- income (Li et al., 2013; Yereli et al., 2014; Falconi & Bernabé, 2018; Tokatlıoğlu & Tokatlıoğlu, 2018; Akhtar et al., 2020; Thu Thuong et al., 2021),
- age of the household head (Choi et al., 2015; Zhao et al., 2020; Thu Thuong et al., 2021),
- employment status of the household head (Yereli et al., 2014; Rashad & Sharaf, 2015; Cleopatra & Eunice, 2018; Tokatlıoğlu & Tokatlıoğlu, 2018; Thu Thuong et al., 2021)
- insurance status of the household head (Yardim et al., 2010; Li et al. 2013; Yereli et al., 2014; Narcı et al. 2015; Tokatlıoğlu & Tokatlıoğlu ,2018; Akhtar et al., 2020; Dalui et al., 2020; Thu Thuong et al., 2021),
- marital status of the household head (Yereli et al., 2014; Tokatlıoğlu & Tokatlıoğlu, 2018),
- gender of the household head (Choi et al., 2015; Rashad & Sharaf, 2015; Cleopatra & Eunice, 2018; Tokatlıoğlu & Tokatlıoğlu, 2018; Dalui et al., 2020),
- education level of the household head (Li et al., 2013; Yereli et al., 2014; Choi et al., 2015; Narcı et al., 2015; Tokatlıoğlu & Tokatlıoğlu, 2018; Zhao et al., 2020),
- household size (Arsenijevic et al., 2013; Yereli et al., 2014; Narcı et al., 2015; Rashad & Sharaf, 2015; Tokatlıoğlu & Tokatlıoğlu, 2018; Vahedi et al., 2020; Thu Thuong et al., 2021),
- sick/disabled person (Yardim et al., 2010; Arsenijevic et al., 2013; Li et al., 2013; Yereli et al., 2014; Choi et al., 2015; Narcı et al., 2015; Piroozi et al., 2016; Ahmed et al., 2018; Falconi & Bernabé, 2018; Tokatlıoğlu & Tokatlıoğlu, 2018; Zhao et al., 2020; Thu Thuong et al., 2021),
- elderly person (Yardim et al., 2010; Barros et al., 2011; Li et al., 2013; Yereli et al., 2014; Choi et al., 2015; Narcı et al., 2015; Piroozi et al., 2016; Ahmed et al., 2018; Falconi & Bernabé, 2018; Tokatlıoğlu & Tokatlıoğlu, 2018; Zhao et al., 2020; Thu Thuong et al., 2021),
- residential area (Yardim et al., 2010; Arsenijevic et al., 2013; Li et al., 2013; Yereli et al., 2014; Cleopatra & Eunice, 2018; Falconi & Bernabé, 2018; Tokatlıoğlu & Tokatlıoğlu, 2018; Akhtar et al., 2020; Zhao et al., 2020)

- between zero-five years old person (Yardim et al., 2010; Li et al., 2013; Yereli et al., 2014; Narcı et al., 2015; Rashad & Sharaf, 2015; Tokatlıoğlu & Tokatlıoğlu, 2018),
- access to health facilities (Tokatlıoğlu & Tokatlıoğlu, 2018)

3. Material and Method

3.1. Data Source and Variables

The data used in the study belongs to the Household Budget Survey (HBA) which is gathered by Turkish Statistical Institute (TurkStat) regularly each year. Data belonging to 40,033 households between the years 2009-2012 were used in the analysis (TurkStat, 2009; TurkStat 2010; TurkStat, 2011; TurkStat, 2012). The data set includes individual, household and consumption characteristics of all households. The expenditure data used have been adjusted for inflation, and a consistent unity has been achieved between all individual and household characteristics.

When we look at what expenditure items are included in out-of-pocket health expenditures; doctors' consultation fees, diagnosis, treatment, examination, medicine and hospital expenses are included, while special nutrition and health-related transportation expenses are not included. Besides, expenditure on alternative and/or conventional medicine is included in OOPHE (Xu, 2005).

Health expenditure variable is used as the dependent variable in the study; it was obtained by taking the sum of all expenses related to pharmaceutical products, other medical products, therapeutic instruments and equipment, medical services (general practitioner and specialist physician), dental services, paramedical services, and hospital services. The dependent and independent variables used in the study are specified in Table 1;

Number	Name	Description	Values			
1	Catastrophe	Household exposure to catastrophe	0	Not catastrophic	1	Catastrophic
2	HHGen	Household head's gender	1	Male	2	Female
3	HHAge	Household head's age	1	Between the age of 15- 29	3	Between the age of 45- 59
			2	Between the age of 30 - 44	4	60 years old and older
4	HHEdu	Household head's education level	1	Illiterate/unschooled	3	High school
			2	Primary school/Secondary school	4	Graduate
			5	Master/Doctorate		
5	HHMar	Household head's marital status	1	Never married	3	His wife or her husband decedent
			2	Married	4	Divorced
6	HHIns	Household head's insurance status	0	Uninsured	1	Insured
7	HHEmp	Household head' employment status	0	Not working	1	Working
9	OSStatus	Ownership status in the residence	1	Homeowner	3	Lodgement
			2	Tenant	4	Other
10	ResArea	Rural-Urban status	1	Rural	2	Urban
11	HSize	Household size	1	HSize<=5	2	HSize>5

12	Income	Annual disposable income	Continuous			
13	AccessHealth	Access to health institutions	1	Hard	2	Easy
14	ZFYearsOld	Presence between zero to five years old person in the household	0	Not available	1	Available
15	65+YearsOld	Presence 65 years old or over the person in the household	0	Not available	1	Available
16	Sick/DisPer	Presence of a person with a physical or mental problem in the household that hinders daily activity	0	Not available	1	Available
17	HHType	Household type	1	Nuclear family	3	Single adult family
			2	Extended family	4	People living together

Table 1. Dependent and Independent Variables Used in The Study

3.2. Methods

Many definitions have been made about the concept of data mining. In simple terms, Data Mining is the process of discovering new, valuable and important models, summaries and derived values from a specific data collection (Kantardzic, 2019). More clearly, data mining is an application-driven and interdisciplinary domain that combines visualization, machine learning, algorithms, database systems, data warehouses, information retrieval, statistics, pattern recognition, and high-performance computing techniques to extract information from large databases (Han et al., 2011).

Many definitions have been made about the concept of data mining. In simple terms, Data Mining is the process of discovering new, valuable and important models, summaries and derived values from a specific data collection (Kantardzic, 2019). More clearly, data mining is an application-driven and interdisciplinary domain that combines visualization, machine learning, algorithms, database systems, data warehouses, information retrieval, statistics, pattern recognition, and high-performance computing techniques to extract information from large databases (Han et al., 2011).

In data mining there are many methods, such as decision trees, association rules, support vector machines, clustering analysis, artificial neural networks, which are used for different aims and targets. Decision trees is a predictive model, which are often used in medicine, engineering, marketing and finance fields in classification and regression problems (Rokach & Maimon, 2008). Decision trees are particularly attractive in the data mining field for a variety of reasons. First, decision trees are nonparametric methods. Second, they can handle both nominal and numerical input values. Third, they can handle data sets that may have outliers (except target value) and missing values. Fourth, they facilitate the explanation of the model, as it is easy to interpret. Last, they are fast to train (Gehrke, 2003; Rokach & Maimon, 2010; Nisbet et al., 2018). This method includes a set of rules to divide a large collection of heterogeneous records into smaller, more homogeneous groups to a specific target variable. The clusters that emerge with each successive division are increasingly similar to each other (Berry & Linoff, 2004). Decision tree algorithms, in other words inducers, construct a decision tree from a data set automatically. Generally, the aim is to minimize the error of generalization and to acquire the optimal tree. Also, it is possible to define other goal functions, such as minimizing the number of nodes or average depth (Rokach & Maimon, 2015). There are lots of decision tree algorithms which differ in terms of the path they follow in choosing root, node, and branching criteria in the literature (Tapkan et al., 2011). These algorithms are AID, THAID, CHAID, CART, ID3, C4.5, See5/C5.0 and QUEST (Sutton, 2005).

CHAID method was used in the study. Chi-square automatic interaction detector shortly CHAID was developed by Kass in 1980 (Kass, 1980). In the CHAID analysis, the best split is found for each explanatory variable. Then, the explanatory variables are compared until the best explanatory variable is selected, and repartitions are made according to the best explanatory variable. All subsections are reanalyzed independently (Pehlivan, 2006). It uses the Chi-square test to determine the best splits (at each step) in classification problems, and F-Test in regression problems (if the target variable is continuous). It is fast, creates wider decision trees, and can provide multiple terminal nodes connected to a single branch (Nisbet et al., 2018).

Data Mining processes and methods, which have become widespread and frequently preferred in recent years, have been used to determine household characteristics that affect CHE. It is known that data mining processes and methods are used in various fields such as diagnosis and classification of diseases and determination of factors causing disease in the field of health. However, it is seen that it is not widely used in health expenditures or health economics. The most important point that reveals the article's importance is to try to reach the result with a method different from the studies performed so far.

4. Findings

Descriptive statistics are included in the study. Table 2 shows the descriptive statistics of the household data used in the study.

Characteristics	Description	Number of observation	Percentage
HHIns	Uninsured	3291	8.2
	Insured	36742	91.8
HHInsType	Public or private insurance (except green card)	32528	81.3
	Green card	4214	10.5
	Uninsured	3291	8.2
HHEmp	Not working	12587	31.4
	Working	27446	68.6
HHMar	Never married	1126	2.8
	Married	34065	85.1
	His wife or her husband decedent	3716	9.3
	Divorced	1126	2.8
HHGen	Male	34524	86.2
	Female	5509	13.8
HHAge	Between the age of 15- 29	2885	7.2
	Between the age of 30 - 44	14418	36.0
	Between the age of 145- 59	13541	33.8
	60 years old and older	9189	23.0
HHEdu	Illiterate/unschooled	5398	13.5
	Primary school and Secondary school	22607	56.5
	High school	6806	17.0

	Graduate	4472	11.9
	Master/Doctorate	450	1.1
HSize	Household size≤5	34620	86.5
	Household size>5	5413	13.5
HHType	Nuclear family	27831	69,5
	Extended family	6431	16,1
	Single adult family	4689	11,7
	People living together	1082	2,7
AccessHealth	Hard	10954	27.4
	Easy	29079	72.6
ZFYerarsOld	Not available	29234	73.0
	Available	10799	27.0
65+YearsOld	Not available	31460	78.6
	Available	8573	21.4
Sick/DisPer	Not available	33655	84.1
	Available	6378	15.9
ResArea	Rural	12567	31.4
	Urban	27466	68.6
Ownership status in the residence	Home owner	24605	61.5
	Tenant	8947	22.4
	Lodgment	924	2.3
	Other	5557	13.9

Table 2. Descriptive Statistics

Table 3 shows the number of households with CHE and OOPHE. While 247 out of 40033 households made CHE, it was observed that 25103 households made OOPHE, which corresponds to 0.62%, and 62.71%, respectively.

Type of Status	Description	Number of observation	Percentage
CHE	0 (Not CHE)	39786	99.38
	1 (CHE)	247	0.62
OOPHE	1 (Made OOPHE)	25103	62.71
	0 (No OOPHE)	14930	37.29

Table 3. Household exposed to CHE and OOPHE

After the descriptive statistics, it came to interpreting the data mining model established to determine the variables that cause a catastrophe. As a result of the established model, the CHAID algorithm's result is shown in Figures 1, 2, 3, and 4. Since it is impossible to display the resulting Decision Tree on a single page, the Decision Tree is divided into four figures.

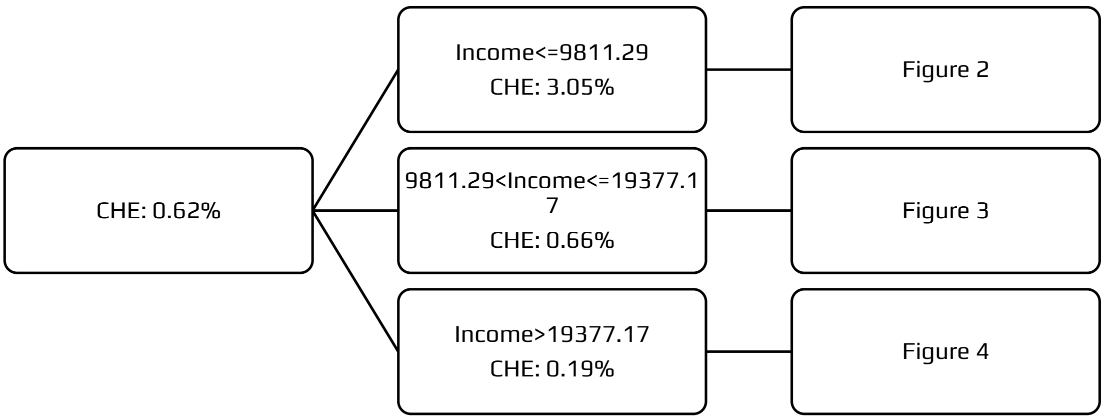

Figure 1. General Outlook of the CHAID Tree

When the Decision Tree is interpreted, the following extractions have been reached. It is seen that the first variable that splits the tree is income. Income emerges as the most important distinguishing characteristic used in determining whether a household is catastrophic or not. When we look at income values in the resulting tree structure, it would not be wrong to divide this variable into three categories as low, medium, and high.

The proportion of households that have been exposed to CHE in the node where the household's income is less than 9811 TL is 3.05%, 0.66% in the node where the household's income between 9811 TL and 19377 TL, and in the node where the income is over 19377 TL as 0.19% is observed. According to these results, it can be said that as the annual disposable income of the household increases, the probability of exposure to CHE decreases and vice versa.

In the node where the household income is 9811 TL and below, exposure to the households' catastrophe situation is shown in Figure 2. Accordingly, the first important characteristic that performs the split in this node is seen as sick/disabled individuals in the household. In the node where there are no sick/disabled individuals, the household proportion exposed to CHE is 1.92%, in the node where sick/disabled person in the household is available this proportion is 6.31%. In households with low-income levels, the possibility of exposure to CHE increases if there are sick/disabled people.

In the node where the income level is low, there is no sick/disabled person, the residential area is seen as the distinctive characteristic that enables split. In the node where there is low-income, where there is no sick/disabled person, if the residence place is rural, the proportion of households exposed to CHE is 2.35%, while in the node where the residence place is urban this proportion is 1.23%. In low-income households, having a sick/disabled individual increases the household's probability of being exposed to catastrophe.

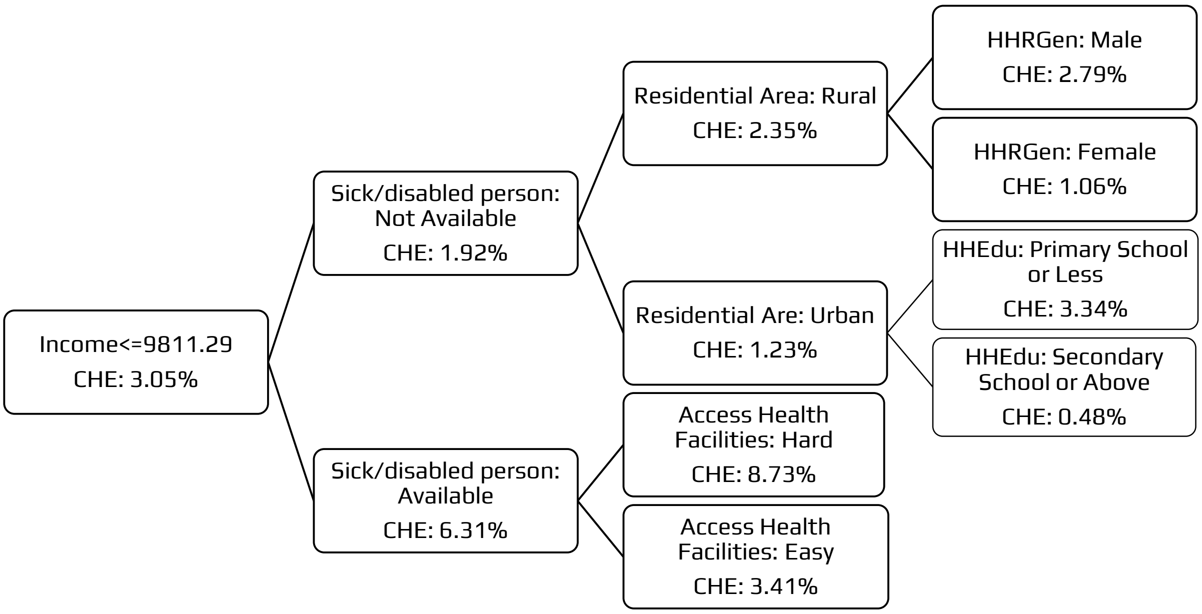

Figure 2. The Outlook of Low-Income Households to Exposure to CHE

In the node where the household income level is low, where there is no sick/disabled individual in the household, the place of the residence is rural, the household head's gender is seen as a distinctive characteristic. In the node where the household head's gender is male, the proportion of households exposed to CHE is 2.79%, while it is 1.06% in the node where the household head's gender is female.

In the node where the household income level is low, there is a sick/disabled indvidual in the household, where households live in the urban, the household head's education level is seen as a distinctive characteristic. The proportion of households exposed to CHE in the node where the household head's education level is primary education or less is 3.34%, while it is 0.48% in the node where the household head's education level is a high school or above.

In the node where the household income level is low, where the household with sick/disabled individual, it is the characteristic of access to healthcare facilities that enable the split. In the node where the households with difficult access to healthcare facilities, the proportion of households exposed to CHE is 8.73%, while this proportion is 3.41% in the node with easy access to healthcare facilities. In the node where the household income level is low, where the household has a sick/disabled individual, where it is difficult to access to healthcare facilities, the proportion of exposure to CHE is the node with the highest proportion among all nodes.

In the node where the household income is between 9811 and 19377 TL, exposure to CHE situation of the households is shown in Figure 3. The first important characteristic that performs the split in this node is the household size. In the node where the household size is five individuals or less than 5, the proportion of the households exposed to CHE is 0.50%, while it is 1.65% in the node with more than five individuals. In the middle-income group, as the number of individuals in the household increases, the probability of household's exposure to CHE increases and vice versa.

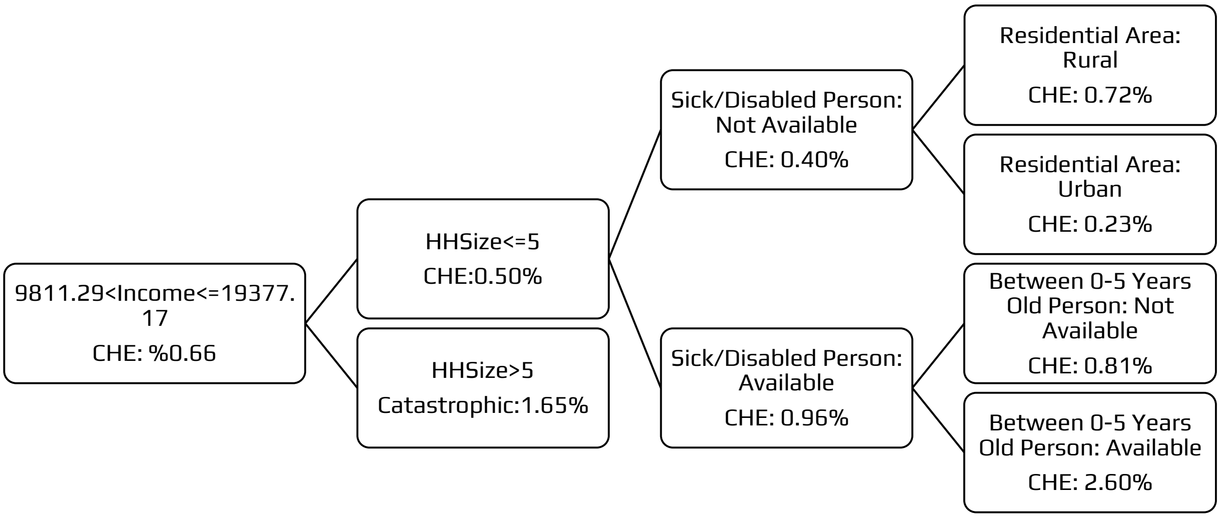

Figure 3. The Outlook of Middle-Income Households to Exposure to CHE

In the node where the household income is between 9811 and 19377 TL, where the household size is five individuals or less than 5, the presence of a sick/disabled individual in the household is seen as a distinctive characteristic. In the node where there is a presence of sick/disabled individuals, the proportion of the household’s exposure to CHE is 0.96%, in the other node, this proportion is 0.40%.

In the node where the household income is between 9811 and 19377 TL, where the number of individuals is five or less than five and does not have sick/disabled individuals, the residential area is distinctive. In the node where the household lives in the rural area, the proportion of household exposure to CHE is 0.72%, while in the node where the household lives in an urban area this proportion is 0.23%.

In the node where the household income is between 9811 and 19377 TL, where the number of individuals is five or less than 5, where there is a presence of sick/disabled individuals, the presence of individuals aged between 0-5 years old is seen as distinctive characteristic. In the node where the households have 0-5-year-old individuals, the proportion of household's exposure to che is 2.60%, while this proportion is 0.81% in the other node.

In the node where the household income is between 9811 and 19377 TL and household size is above 5 individuals, no distinctive characteristic has been observed. more clearly, no splitting has occurred in this node

In the node where the household income is above 19377 TL, exposure to the households' catastrophe situation is shown in figure 4. in this node with high-income households, the proportion of households exposed to CHE is 0.19%. in this node, the first important characteristic that causes the split is whether there are individuals aged 65+ in the household. ın the node where the is a presence of 65+ individuals, the proportion of household's exposure to CHE is 0.65%, while in the node where there are no individuals aged 65+ in the household this proportion is 0.10%.

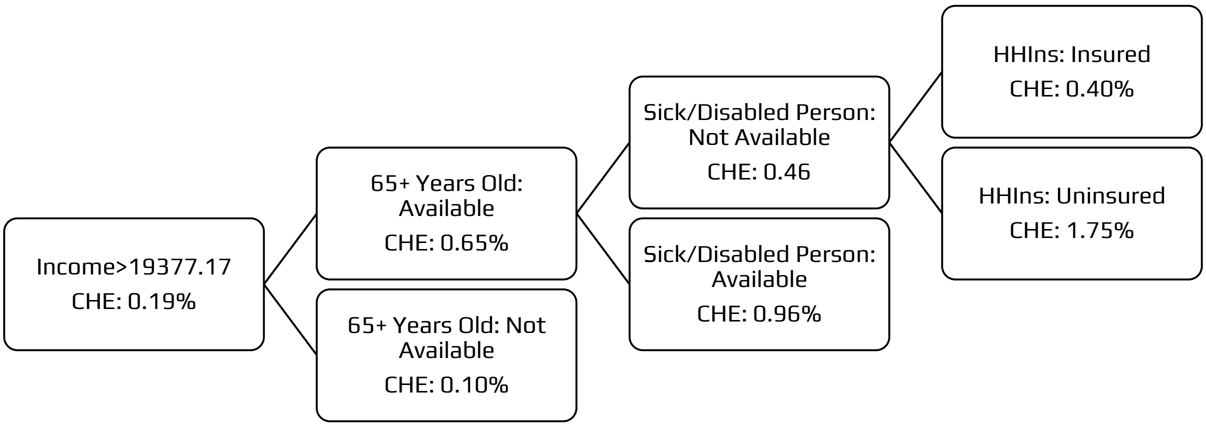

Figure 4. The Outlook of High-Income Households to Exposure to CHE

In the node where the household income is over 19377 TL, where there are individuals aged 65+ in the household, the presence of sick/disabled individuals in the household is seen as the distinctive characteristic. In the node where there are sick/disabled individuals, the proportion of households exposure to CHE is 0.96%, while in the other node this proportion is 0.46%.

In the node where the household income is above 19377 TL, where there are individuals aged 65+, where there are no sick/disabled individuals, the head of the household's insurance status is seen as the distinctive characteristic. In the node where the head of the household is insured, the proportion of the households exposure to CHE is 0.40%, while it is 1.75% in the node where the head of the household is uninsured.

In the node where household income is over 19377 TL, where there are individuals aged 65+, where there is a sick/disabled individual, the proportion of households exposure to CHE is 0.96%. No distinctive characteristics have been observed in this node; therefore, there is no splitting occurred.

5. Discussion

According to the results of the CHAID analysis carried out; income, presence of a person sick/disabled in the household, residential area, household size, education level, gender and health insurance status of the head of the household, presence of individuals aged 65+, presence between zero to five years old people, and access to health institutions characteristics have been observed to affect CHE. There was no effect of the characteristics of employment status and marital status of the head of the household, household type, and ownership status in residence on CHE.

In the decision tree, it has been observed that the first distinguishing variable is income, we can specify the 3 sub-branches that occur as low, medium, and high-income groups. In the low-income group, it is observed that the presence of a sick/disabled person in the household, residential area, the access to health facilities, and the gender and education level of the household head affect the CHE. The characteristics that affect catastrophe in the middle-income group are the size of the household, the residential area, the presence of a sick/disabled person, and the presence of individuals between the age of zero and five in the household. In the high-income group, it has been observed that the presence of an individual aged 65+, the insurance status of the household head, and the presence of sick/disabled person have effects on CHE. It has been observed that the presence of a sick/disabled person affects the CHE at all income levels.

When we examine each variable that affects CHE separately, income is the most important variable. Several studies have indicated that income affects CHE (Li et al., 2013; Falconi & Bernabé, 2018; Akhtar et al., 2020; Thu Thuong et al., 2021). Obtained from CHAID analysis, the results support studies conducted before in Turkey (Yereli et al., 2014; Tokatlıoğlu & Tokatlıoğlu, 2018). It was observed that the proportion of households exposed to catastrophe decreases as the income of the household increases and vice versa.

In most of the studies (in all of the studies conducted in Turkey) examined, it was stated that the presence of sick/disabled individuals in the household affetcs the CHE (Arsenijevic et al., 2013; Choi et al., 2015; Narcı et al., 2015; Piroozi et al., 2016; Ahmed et al., 2018; Zhao et al., 2020). Because sick/disabled individuals may need health and care services frequently and intensely, these households are likely to be exposed to CHE. The results obtained in this study support the findings made in Turkey and other countries. It has been observed that the presence of a sick/disabled person in the household affects catastrophe at all income levels. After the household income, it is the most important factor affecting CHE. According to the CHAID analysis result, households with sick/disabled individuals are more likely to be exposed to CHE than households without these individuals.

It is expected that households where 65+ individuals live are more likely to be exposed to CHE than other households, as the individuals aged 65+ may often need doctors, medicines and care services. In most of the studies (in all of the studies conducted in Turkey), elderly members in the household affect CHE (Yardim et al., 2010; Yereli et al., 2014; Narcı et al., 2015; Ahmed et al., 2018; Tokatlıoğlu & Tokatlıoğlu, 2018; Si et al., 2019; Thu Thuong et al., 2021). According to the results, an elderly person in the household increases the probability of household's exposure to CHE. In particular, diseases that occur in older ages may be the cause of this situation. While in the upper-income group, it is observed that the presence of a 65+ year-old person in the household affects the CHE, it is the most important characteristic that enables the split in this income group.

As preschool children may need more intensive health services, it is expected that households where these individuals live are more likely to be exposed to CHE than other households. Some studies (Li et al., 2013; Rashad & Sharaf, 2015) stated that the presence of a person aged between 0-5 in the household increases the household's probability of exposure to CHE, while some studies, such as Yardim et al. (2010), state otherwise. According to our results, in middle-income households, the presence of individuals between the ages of 0-5 increases the probability of household's being exposed to CHE.

The gender of the head of the household variable is seen as an effective variable in the low-income group. Some studies report that the gender of the household head affects CHE (Choi et al., 2015; Cleopatra & Eunice, 2018; Dalui et al., 2020). Rashad and Sharaf (2015) stated that if the household head is female, the probability of the household being exposed to CHE decreases, while Cleopatra and Eunice (2018) and Dalui et al. (2020) stated the opposite. The results obtained from CHAID analysis show the exact opposite of the study of Tokatlıoğlu & Tokatlıoğlu (2018) conducted in Turkey before. In the low-income group, if the household head is male, the household's probability of being exposed to catastrophe increases.

In several studies mentioned in the literature section, it is stated that the education level of the household head affects catastrophe (Li et al., 2013; Zhao et al., 2020). Li et al. (2013) and Tokatlıoğlu & Tokatlıoğlu (2018) stated that if the education level of the household head is high, the probability of the household being exposed to CHE is low. Individuals are expected to have a better job as their education level rises, and as a result, they will have a higher-income job. It is expected that exposure to CHE in a household with a high income will be low. According to the analysis results, the household head's education level affects the catastrophe in low-income groups. As the household head's education level increases in low-income groups, the proportion of households exposure to catastrophe decreases.

Households with health insurance are expected to be less likely to be exposed to CHE due to their lower health expenditure and vice versa. Many studies stated the insurance status of the household head affects CHE (Li et al., 2013; Akhtar et al., 2020; Dalui et al., 2020; Thu Thuong et al., 2021). Also, in all of the studies carried out in Turkey, insurance status of the head of the household is stated to affect the CHE (Yardim et al., 2010; Yereli et al., 2014; Narcı et al., 2015; Tokatlıoğlu & Tokatlıoğlu, 2018). According to our study results, the insurance status of the head of the household affects the catastrophe too. The uninsured household head was more likely to be exposed to catastrophe than the insured household head.

Again, it was stated in some studies that household size, which is another important characteristic, affects CHE (Arsenijevic et al., 2013; Rashad & Sharaf, 2015; Vahedi et al., 2020; Thu Thuong et al., 2021). As the household size increases, income can be expected to increase due to the increase of individuals who bring income to the household. Alternatively, with the increase in the number of individuals, household expenses can be expected to increase. Therefore, the effect of this variable can take different forms. Looking at the studies conducted in Turkey, Yereli et al. (2014) and Tokatlıoğlu & Tokatlıoğlu (2018), the probability of CHE decreases as the household size increases, Yardim et al. (2014) found no effect of household size on catastrophe. Interestingly, according to our analysis results, in the middle-income group, as the household's size increases, the probability of the household's exposure to CHE increases.

Another important characteristic is the residential area. Due to household's being located in rural areas may cause them to spend more to access health services. Hence, households are expected to be more likely to be exposed to CHE. Looking at the impact of the residential area on the CHE, many studies find that households in rural residential areas are more likely to be exposed to CHE than urban households (Cleopatra & Eunice, 2018; Falconi & Bernabé, 2018; Akhtar et al., 2020). Our study supports earlier studies carried out in Turkey (Yardim et al., 2010; Yereli et al., 2014). Both in the low income and middle-income groups, rural households are more likely to be exposed to CHE than those living in urban areas.

Access to health facilities of the household is another characteristic that affects CHE. Tokatlıoğlu & Tokatlıoğlu (2018)'s study has indicated that access to healthcare facilities affect CHE. According to the results of our study, it is seen that this variable affects CHE. These findings support the Tokatlıoğlu & Tokatlıoğlu (2018)'s research conducted in Turkey before. In the low-income group, with the households living in rural areas where access to healthcare facilities are hard, they are more likely to be exposed to che than the households that have easy access to healtcare facilities.

Some studies stated that the age of the household head affect CHE (Zhao et al., 2020; Thu Thuong et al., 2021). However, according to the study results, no effect of this characteristic on CHE was observed.

Yereli et al. (2014) and Tokatlıoğlu & Tokatlıoğlu (2018) stated that the marital status of the household head affects CHE. But according to the CHAID analysis no effect of the marital status of the household head on CHE was observed.

Some studies stated that the household head's employment status affects CHE (Cleopatra & Eunice, 2018; Thu Thuong et al., 2021). Also, some studies conducted in Turkey stated that the employment status of the household head affects CHE (Yereli et al., 2014; Tokatlıoğlu & Tokatlıoğlu, 2018). According to the Narcı et al. (2014) study conducted in Turkey, the household head's employment status has no impact on CHE. According to the CHAID analysis results, the household head's employment status does not affect CHE. Finally, according to the study results, no effect of household type on CHE was observed.

6. Conclusion

In the study conducted to determine the factors affecting households' CHE, CHAID analysis, which is a method different from econometric or statistical methods, was used. According to the results of the CHAID analysis carried out; household income, presence of a sick/disabled person in the household, the rural/urban residential area status, household size, education level, gender and insurance status of the head of the household, the presence of aged 65+ person in the household, the presence of a person between the ages of zero and five in the household, access to health facilities of the household affected catastrophe. The marital status, age and employment status of the household head, ownership status in residence and the household type have no effect on CHE.

It was especially observed that households with low income, with sick/disabled individuals, and those with difficult access to healthcare facilities are more likely to be exposed to CHE. The reason why these households are more likely to be exposed to CHE can be the fact that the sick/disabled individuals need more health services, the household spends more on health due to the distance from health institutions, and the household income is low. The high probability of exposure to CHE is seen in nodes with low-income households. In the node of middle-income households, it was observed that households where the household size is less than 6, where there is a sick/disabled individual, and where there is an individual aged 0-5 were most likely to be exposed to CHE. In the node of high-income households, it was observed that households with individuals over the age of 65+, where there are no sick/disabled individuals, and where there are uninsured household heads, were most likely to be exposed to CHE.

The residential area variable was seen as a distinguishing feature in nodes with low and middle-income households, and rural residents were more likely to be exposed to CHE in both income groups. For this reason, it may be considered to locate more health facilities in number in rural areas so that rural households can easily access them.

The only characteristic seen at all income levels is whether there is a sick/disabled person in the household. Since households with disabled/sick individuals are more likely to be exposed to CHE, providing, especially low-income households, with more in-kind or cash assistance may be a method to prevent CHE to a certain extent.

None of the households exposed to CHE have supplementary health insurance. Taking out supplementary health insurance can prevent households from being exposed to CHE.

The proportion of households exposed to CHE was 0.62%, and the proportion of households making health expenditures was 62.71%. Results of the CHAID analysis supports earlier studies conducted in Turkey (Yardim et al., 2010; Yereli et al., 2014; Narcı et al., 2015; Tokatlıoğlu & Tokatlıoğlu, 2018). In addition, the results obtained support the studies mentioned in the literature section. Details about these are explained extensively in the discussion section of the article. I believe that data mining methods, which are frequently used in health, are an important contribution in terms of being used for the first time in determining the determinants of catastrophic health expenditures. In addition, this study can guide and give ideas to those working in health economics about using data mining methods.

References

Ahmed, S., Szabo, S., & Nilsen, K. (2018). Catastrophic healthcare expenditure and impoverishment in tropical deltas: Evidence from the Mekong Delta region. International Journal for Equity in Health, 17:53, 1–13.

Akhtar, A., Ahmad, N., & Chowdhury, I. R. (2020). Socio-economic inequality in catastrophic health expenditure among households in India: A decomposition analysis. Indian Economic Review, 55, 339-369.

Arsenijevic, J., Pavlova, M., & Groot, W. (2013). Measuring the catastrophic and impoverishing effect of household health care spending in Serbia. Social Science & Medicine, 78, 17–25.

Barros, A. J. D., Bastos, J. L., & Dâmaso, A. H. (2011). Catastrophic spending on health care in Brazil: Private health insurance does not seem to be the solution. Cad. Saude Publica, 27, 254–262.

Berry, M. J. A. and Linoff, G. S. (2004). Data Mining Techniques For Marketing, Sales, and Customer Relationship Management (Second Edition). Indianapolis, Indiana: Wiley Publishing Inc.

Choi, J. W., Choi, J. W., Kim, J. H., Yoo, K. B., & Park, E. C. (2015). Association between chronic disease and catastrophic health expenditure in Korea. BMC Health Services Research, 15:26, 1–8.

Cleopatra, I., & Eunice, K. (2018). Household Catastrophic Health Expenditure : Evidence from Nigeria. Microeconomics and Macroeconomics, 6(1), 1–8.

Dalui, A., Banerjee, S., & Roy, R. (2020). Determinants of out-of-pocket and catastrophic health expenditure in rural population: A community-based study in a block of Purba Barddhaman, West Bengal. Indian Journal of Public Health, 64, 223–228

Falconi, D. P., & Bernabé, E. (2018). Determinants of catastrophic healthcare expenditure in Peru. International Journal of Health Economics and Management, 18, 425–436.

Gehrke, J. (2003). Decision Trees, (Ed.) Ye, N. In The Handbook of Data Mining (1st Edition). New Jersey: CRC Press.

Goryakin, Y., & Suhrcke, M. (2014). The prevalence and determinants of catastrophic health expenditures attributable to non-communicable diseases in low- and middle-income countries : a methodological commentary. International Journal for Equity in Health, 13:107, 1–5.

Han, J., Kamber, M., & Pei, J. (2011). Data Mining Concepts and Techniques Third Edition. The Morgan Kaufmann Series In Data Management Systems. USA, Waltham: Elsevier.

Kang, J. H., & Kim, C. W. (2018). Relationship Between Catastrophic Health Expenditures and Income Quintile Decline, Osong Public Health and Research Perspectives, 9(2), 73–80.

Kantardzic, M. (2019). Data Mining Concepts, Models, Methods, and Algorithms Third Edition. New Jersey: John Wiley & Sons, Inc.

Kass, G. V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. Applied Statistics, 29(2), 119-127.

Knaul, F. M., Wong, R., Arreola-Ornelas, H., Méndez, O., & The Network on Health Financing and Social Protection in Latin America and the Caribbean (LANET) (2011). Household catastrophic health expenditures: a comparative analysis of twelve Latin American and Caribbean Countries. Salud Pública de México, 53(2), 85–95.

Li, X., Shen, J. J., Lu, J., Wang, Y., Sun, M., Li, C., Chang, F., & Hao, M. (2013). Household catastrophic medical expenses in eastern China: Determinants and policy implications. BMC Health Services Research, 13:506, 1-9.

Memişoğlu, D., & Durgun, A. (2011). 2008 Ekonomik Krizi ve Türkiye'de Sağlık Sektörü Üzerine Etkileri. Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 1(13), 81–100.

Narcı, H. Ö., Şahin, İ., & Yıldırım, H. H. (2015). Financial catastrophe and poverty impacts of out-of-pocket health payments in Turkey. The European Journal of Health Economics, 16(3), 255–270.

Nisbet, R., Miner, G., & Yale, K. (2018). Handbook of Statistical Analysis and Data Mining Applications (Second Edition). London: Academic Press, Elsevier Inc.

Njagi, P., Arsenijevic, J., & Groot, W. (2018). Understanding variations in catastrophic health expenditure, its underlying determinants and impoverishment in Sub-Saharan African countries: a scoping review. Systematic Reviews, 7:136, 1-23.

O'Donnell, O., van Doorslaer, E., Wagstaff, A., & Lindelow, M. (2008). Analyzing Health Equity Using Household Survey Data A Guide to Techniques and Their Implementation, Washington, DC: The World Bank.

OECD (2019). Health at a Glance 2019: OECD Indicators. OECD Publishing, Paris.

Pehlivan, G. (2006). CHAID Analizi ve Bir Uygulama, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.

Piroozi, B., Moradi, G., Nouri, B., Bolbanabad, A. M., & Safari, H. (2016). Catastrophic health expenditure after the implementation of health sector evolution plan: A case study in the west of Iran. International Journal of Health Policy and Management, 5(7), 417–423.

Sutton, C. D., (2005). Classification and Regression Trees, Bagging, and Boosting, (Ed.) Rao, C. R., Wegman, E. J., & Solka, J. L. In Data Mining and Data Visualization, Handbook of Statistics, Vol. 24, (pp. 303-329). The Netherlands: Elsevier B.V.

Rashad, A. S., & Sharaf, M. F. (2015). Catastrophic and impoverishing effects of out-of-pocket health expenditure: New evidence from Egypt. American Journal of Economics, 5(5), 526–533.

Rokach, L., & Maimon, O. (2008). Data Mining With Decision Trees Theory and Applications. Singapore: World Scientific Publishing Co. Pte. Ltd.

Rokach, L., & Maimon, O. (2010). Classification Trees. (Ed.) Maimon, O. & Rokach, L., In Data Mining and Knowledge Discovery Handbook Second Edition, (pp. 149-174), New York: Springer, 149-174.

Rokach, L., & Maimon, O. (2015). Data Mining With Decision Trees Theory and Applications (2nd Edition). Singapore: World Scientific Publishing Co. Pte. Ltd.

Si, Y., Zhou, Z., Su, M., Wang, X., Lan, X., Wang, D., Gong, S., Xiao, X., Shen, C., Ren, Y., Zhao, D., Hong, Z., Bian, Y., & Chen, X. (2019). Decomposing inequality in catastrophic health expenditure for self-reported hypertension household in Urban Shaanxi, China from 2008 to 2013: Two waves' cross-sectional study. BMJ Open, 9, 1–9.

T.C. Maliye Bakanlığı, (2010). Merkezi Yönetim Kapsamındaki İdareler, Sosyal Güvenlik Kurumları ve Mahalli İdareler 2009 Yılı Genel Faaliyet Raporu, Ankara.

Tapkan, P., Özbakır, L., & Baykasoğlu, A. (2011). Weka İle Veri Madenciliği Süreci ve Örnek Uygulama. Endüstri Mühendisliği Yazılımları ve Uygulamaları Kongresi, 30 Eylül-02 Ekim 2011, İzmir. (pp. 247–262).

Thu Thuong, N. T. , Berg, Y. V. D., Huy, T. Q., Tai, D. A., & Anh, B. N. H. (2021). Determinants of Catastrophic Health Expenditure in Vietnam. The International Journal of Health Planning and Management, 36, 316–333.

Tokatlıoğlu, Y., & Tokatlıoğlu, İ. (2018). Türkiye'de Katastrofik Sağlık Harcamaları ve Bu Harcamaları Belirleyen Faktörler: 2002-2014 Dönemi. Sosyoekonomi, 26(35), 59–78.

TurkStat, (2009). Turkey Household Budget Survey-2009, Ankara.

TurkStat, (2010). Turkey Household Budget Survey-2010, Ankara.

TurkStat, (2011). Turkey Household Budget Survey-2011, Ankara.

TurkStat, (2012). Turkey Household Budget Survey-2012, Ankara.

UN, (2015). United Nations, Transforming Our World: The 2030 Agenda For Sustainable Development, A/RES/70/1.

Vahedi, S., Rezapour, A., Khiavi, F. F., Esmaeilzadeh, F., Javan-Noughabi, J., Almasiankia, A., & Ghanbari, A. (2020). Decomposition of Socioeconomic Inequality in Catastrophic Health Expenditure: An Evidence from Iran. Clinical Epidemiology and Global Health, 8, 437–441.

Wagstaff, A., Flores, G., Hsu, J., Smitz, M. F., Chepynoga, K., Buisman, L. R., van Wilgenburg, K., & Eozenou, P. (2018). Progress on catastrophic health spending in 133 countries: a retrospective observational study. The Lancet Global Health, 6(2), 169–179.

Wagstaff, A., & van Doorslaer, E. (2003). Catastrophe and impoverishment in paying for health care: with applications to Vietnam 1993-1998. Health Economics, 12, 921–934.

WHO. (2019). Global Spending on Health: A World in Transition 2019. Geneva: World Health Organization

WHO, & UNICEF. (2018). A vision for primary health care in the 21st century: towards universal health coverage and the Sustainable Development Goals. Geneva: World Health Organization and the United Nations Children's Fund (UNICEF).

World Bank. (2021a). Out-of-pocket (health) expenditure per capita (current US$). https://data.worldbank.org (access date: 18.08.2020).

World Bank. (2021b). Current health expenditure per capita (current US$) https://data.worldbank.org (access date: 18.08.2020).

Xu, K., Evans, D. B., Carrin, G., Aguilar-Rivera, A. M., Musgrove, P., & Evans, T. (2007). Protecting households from catastrophic health spending. Health Affairs, 26(4), 972–983.

Xu, K. (2005). Distribution of health payments and catastrophic expenditures methodology. Geneva: World Health Organization, Discussion Paper, Number 2, EIP/HSF/DP.05.2.

Xu, K., Evans, D. B., Carrin, G., & Aguilar-Rivera, A. M. (2005). Designing Health Financing Systems to Reduce Catastrophic Health Expendıtiure, Geneva: World Health Organization, Technical Briefs for Policy-Makers, Number 2, WHO/EIP/HSF/PB/05.02.

Xu, K., Evans, D. B., Kawabata, K., Zeramdini, R., Klavus, J., & Murray, C. J. L. (2003). Household catastrophic health expenditure: a multicountry analysis. The Lancet, 362, 111–117.

Yardim, M. S., Cilingiroglu, N., & Yardim, N. (2010). Catastrophic health expenditure and impoverishment in Turkey. Health Policy, 94, 26–33.

Yereli, A. B., Köktaş, A. M., & Selçuk, I. Ş. (2014). Türkiye'de Katastrofik Sağlık Harcamalarını Etkileyen Faktörler. Sosyoekonomi, 2, 273–296.

Zhao, Y., Oldenburg, B., Mahal, A., Lin, Y., Tang, S., & Liu, X. (2020). Trends and socio-economic disparities in catastrophic health expenditure and health impoverishment in China: 2010 to 2016. Tropical Medicine and International Health, 25(2), 236–247.