

# INFLUENCE OF DEMOGRAPHIC DETERMINANTS ON THE NUMBER OF DEATHS CAUSED BY CIRCULATORY SYSTEM DISEASES IN COMPARISON TO THE NUMBER OF DEATHS CAUSED BY NEOPLASMS IN SLOVAK REGIONS FROM 1996–2014

Ján Fedáčko<sup>1</sup>, Daniel Pella<sup>1</sup>, Beáta Gavurová<sup>2</sup>, Samuel Koróny<sup>3</sup>

<sup>1</sup>1st Department of Internal Medicine, Louis Pasteur University Hospital, Pavol Jozef Šafárik University in Košice, Košice, Slovak Republic

<sup>2</sup>Faculty of Economics, Technical University of Košice, Košice, Slovak Republic

<sup>3</sup>Research and Innovation Centre, Faculty of Economics, Matej Bel University, Banská Bystrica, Slovak Republic

## SUMMARY

**Objectives:** The objective of our study was to evaluate the influence of available demographic determinants on the number of deaths caused by circulatory system diseases as compared to deaths caused by neoplasms in Slovakia in 1996–2014.

**Methods:** Mortality data were kindly provided by the National Health Information Centre in Slovakia. The first method was trend curve fitting of death ratios caused by circulatory system diseases (Chapter IX) and of deaths caused by neoplasms (Chapter II) as a function of age for both sexes. The second method comprised a decision tree for classification between deaths caused by Chapter IX and Chapter II diseases. Input variables were available demographic indicators: age, sex, marital status, region, and calendar year of death. Statistical data analyses were performed by IBM SPSS version 19 statistical software.

**Results:** We found that the odds ratios of deaths caused by circulatory system diseases (Chapter IX) in comparison with deaths caused by neoplasms (Chapter II) were non-decreasing. At first, the values of odds ratios are constant until they reach a critical sex-dependent value with a subsequent steady increase. In the case of men the odds ratio was greater than in the 60 years age-group where the odds ratio value increased slowly (from 1.14 at age 60 to 7.25 at age 90 years). The relative increase was 6.36 (7.25/1.14). The odds ratio in the women group was smaller but increased more rapidly (from 0.81 at age 60 to 12.27 at age 90 years). The relative increase was 15.15 in women (12.27/0.81). Hence, the odds ratio of death caused by Chapter IX diseases vs. Chapter II was greater in the older women group (i.e. higher age values). Utilizing the decision tree model, we have found that the most significant demographic determinant of death counts in both ICD Chapters was the age of the deceased, followed by marital status and finally gender. The last two predictors (year and region) were relatively negligible though formally significant.

**Conclusions:** The proposed method could be useful for prognostic classification of patients and primarily beneficial for hospitals in human or financial resources planning.

**Key words:** mortality, diseases of the circulatory system, neoplasms, decision tree

**Address for correspondence:** D. Pella, 1st Department of Internal Medicine, Louis Pasteur University Hospital, Pavol Jozef Šafárik University in Košice, Trieda SNP 1, 041 90 Košice, Slovak Republic. E-mail: pellad@stonline.sk

<https://doi.org/10.21101/cejph.a5053>

## INTRODUCTION

Differences in life expectancy and in premature mortality due to oncological and cardiovascular diseases between Slovakia and the European Union cannot be explained solely by differences in health care quality. Firstly, one must realize that in all former communist countries, including Slovakia, sufficient financial resources were not invested into costly technologies, medicine and devices needed for early diagnosis and treatment of cancer and cardiovascular diseases. With regards to prevention of aforementioned diseases as well as improving the health status, several influences are crucial. Most diseases arise when the relationship between the human body and the environment in which one lives

is disturbed. One's lifestyle is generally regarded as a decisive factor in relation to the health status. Health-related behaviours, such as eating habits, physical activity, smoking and alcohol consumption along with the prevalence of risk factors such as high blood pressure, high cholesterol or high body weight, significantly influence premature mortality caused mainly by cardiovascular and oncological diseases.

From all recorded deaths in Slovak Republic, roughly three fourths (exact proportion is 76.7%) are caused either by circulatory system diseases (Chapter IX; proportion 53.8%) or by neoplasms (Chapter II; proportion 22.9%) during the years 1996–2014. Hence, it is of critical importance to focus clinical and research interest on diseases of previously outlined chapters.

---

Main objective of our paper was to evaluate the influence of available demographic determinants to deaths caused by the circulatory system diseases (ICD – International Statistical Classification of Diseases and Related Health Problems – I00-I99 – Chapter IX) in comparison to deaths caused by neoplasms (Chapter II, C00-D48). Two different approaches were utilized to achieve the objective. The first approach was trend curve fitting of numbers of deaths as function of the age for both sexes. The second approach was suitable data mining model. Furthermore, we wanted our model to be precise, simple and easy to grasp. That is why we chose decision trees (originally classification trees or regression trees). The decision trees explore relationships between relatively homogenous subgroups and predict future events in an intuitive manner, based on categorical results.

Decision trees have been used widely also in the field of medicine during last years. The application level may have predictive or classification character, which makes this method important in the decision-making processes (medical, socio-economic, etc.). Numerous foreign studies present applications of decision trees also in the other fields of research.

## Literature Review

Due to the objective of the paper, our attention was focused mainly on studies concerned with cancer and cardiovascular diseases. Firstly, Ibrahim et al. (1) examined the survival probability risk factors for breast cancer patients. The results showed that the proposed method had significant benefits for the prognostic classification. Secondly, Dudoit et al. (2) used decision trees to classify tumors based on gene expression, because successful treatment is, for the most part, determined by precise classification of tumors. Moreover, other authors (3, 4) used the method to identify groups of patients for whom the treatment had a positive effect, or to identify risks affecting the probability of developing secondary neoplasms.

Application of models based on outputs of decision trees is very large, determined by a wide range of studies examined. Spitz et al. (5) used decision trees to create suitable risk models to predict lung cancer. Etzel et al. (6) focused on the design and verification of the risk prediction model in lung cancer, specific for African-American population. Tsien et al. (7) used the model approaches in applications of decision trees for the diagnosis and prevention of heart attacks. Furthermore, the previously mentioned authors explicitly emphasized the increasing importance of IT technologies in hospitals, which produces high-quality data base for a wide range of analyses. This was also supported in recent study by Solinger and Rothman (8), which aimed to highlight the important analytical potential in the electronic healthcare records (EMR) and appeal to a broad spectrum of application methods for its use in favor of the improvement of diagnostic and therapeutic processes. According to Solinger and Rothman (8), hospital electronic records can be used in the process of aggregation of all health-related risks with the relevant causal links. Furthermore, specific set of data from various patients is required to explore a wide range of risk factors. Apart from clinical data, genetics are also crucial, as pointed out by Pittman et al. (9). Pittman et al. (9) utilized both genetic and clinical data in a decision trees to predict breast cancer. Gülkesen et al. (10) saw the potential of the decision tree in prediction algorithms for prostate cancer, pointing out the

need to explore new techniques and methods for early diagnosis of this serious disease. Moreover, studies focusing on the relationship between morbidity and mortality in relation to environmental and socio-economic factors should be further explored, as pointed out by Ignatius et al. (11), who utilized decision trees to evaluate waiters. As reported in his study, the waiters are at high risk of cancer, for example nasopharyngeal cancer, and yet this was not examined in any other study. Therefore, Ignatius et al. (11) explicitly aimed to investigate the causal determinants and increased risk of cancer-related mortality, e.g. nasopharyngeal cancer regarding waiters and waitresses in Chinese restaurants. Clary and Ritz (12) examined the relationship between organochlorine pesticides and the incidence of pancreatic cancer. Lee et al. (13) investigated the relationship between the cancer-related mortality and the incidence of chlorinated hydrocarbons in groundwater among people who live near contaminated sites. With regards to socioeconomic factors reflected in decision trees, MacArthur et al. (14) examined the relationship between work and breast-cancer as well as oncological diseases of the reproductive organs.

The previously outlined studies provide interesting cross-section of scientific research studies in the medical field, emphasizing the application of decision trees and their benefits in the implementation of the results into diagnostic and therapeutic practice. Our study aimed to introduce new, innovative mode of application for this method, which will help public health systems to detect the influence of related demographic determinants on the number of deaths caused by circulatory system diseases in comparison to the number of deaths caused by neoplasms.

## MATERIALS AND METHODS

The Vital Registration system routinely collects all the mortality data in Slovakia. Official mortality data have the advantage of being almost totally complete because deaths are unlikely to go unrecorded. Death certificate data include basic demographic information about the decedent as well as information about the cause of death, including the immediate cause and contributing factors.

Data were kindly provided by National Health Information Center in Slovakia. They were in form of classic data matrix with variables in columns and with individual cases in rows. In our paper the data of mortality cases caused by the diseases of the circulatory system and by neoplasms are output variables. Input variables (predictors) are their potential demographic determinants: decedent's age, sex, marital status (single, married, divorced or widowed), Slovak region (NUTS III) and the calendar year of death (range from 1996 to 2014).

The objective of our study was to find out influence of available demographic determinants on number of deaths caused by circulatory system diseases in comparison with number of deaths caused by neoplasms in Slovakia during the years of 1996–2014. So we had to compare mortality counts in both Chapter II and Chapter IX. Later we shall show that the number of death counts was larger in Chapter IX in comparison with death counts in Chapter II for most age values for both sexes. That is why we chose ratio – death counts in Chapter IX/death counts in Chapter II. For curve fitting of mortality curves we needed some basic mortality measures (15, 16). When comparing two countries or

regions in terms of mortalities, the rates are useful if the age population composition is similar. In case of divergent population, the results of comparison might be misleading. Age standardization (direct or indirect) could be a solution to the problem. In our paper, we wanted to show that sometimes such standardization might not be necessary. Our starting point was the definition of mortality odds by Miettinen (17): “Mortality odds is occurrence odds for a given cause of death relative to another.” And meaning of occurrence odds: “for two theoretical rates, having to do with alternative types of state or event in the same base, the ratio of the rate for one type to that for the other”. The last one is mortality odds ratio – “ratio of mortality odds between two categories of a determinant”. This way, the age standardization and related problems could be solved, as will be shown later.

If we divided the number of deaths in Chapter IX by the number of deaths in Chapter II for all age values, then we got “relative risk” of death by circulatory system diseases vs neoplasms. If for example at some particular age the ratio was 2.5, then the probability of death by circulatory system diseases was 2.5 larger than the probability of death caused by neoplasms at this age. The ratio could be expressed by ratio of corresponding crude death rates

$$\frac{CDR_{IX} \frac{D_{IX}}{P}}{CDR_{II} \frac{D_{II}}{P}} = \frac{D_{IX}}{D_{II}} \quad (1),$$

where

$CDR_{IX}$  is crude death rate for Chapter IX,

$CDR_{II}$  is crude death rate for Chapter II,

$D_{IX}$  is number of deaths from Chapter IX,

$D_{II}$  is number of deaths from Chapter II

and  $P$  is midyear population.

Ratio was reduced by  $P$  because it was the same quantity – human population.

Classification methods try to use measured characteristics to divide people or objects into known groups (18). Data mining publications refer to classification as supervised learning, distinguished from cluster analysis as an example of unsupervised learning, in which groups are unknown and must be discovered by analysis. The objective of classification analysis is to find some rule(s) to discriminate among the groups. Example of classification in clinical medicine is problem of differential diagnosis. Observations made on patients are known to fall into specific diagnostic groups (19, 20).

The most common classification methods are: discriminant analysis, logistic regression and decision trees. Both discriminant analysis and logistic regression are special regression parametric methods with results in a form of regression equation. Output values of such regression equation allow individuals to be assigned into a particular group. Contrary to this, decision tree is a new computer data mining, nonparametric method. Decision trees are similar to diagnostic clinical decision making, because they create rules of “if-then” type that are common in clinical practice.

There were several significant associations between mortality odds caused by the diseases of the circulatory system vs neoplasms and available potential influence variables were age, region, sex, marital status and calendar year of death. The objective of our research was to find and present the most significant associations. Therefore decision trees (IBM SPSS version 19) were used.

The decision tree methods available in SPSS are Chi-squared Automatic Interaction Detection (CHAID), Classification and Regression Trees (CART) (20) and Quick, Unbiased, Efficient Statistical Tree (QUEST). In our analysis, CART method was used because its results were relatively the best.

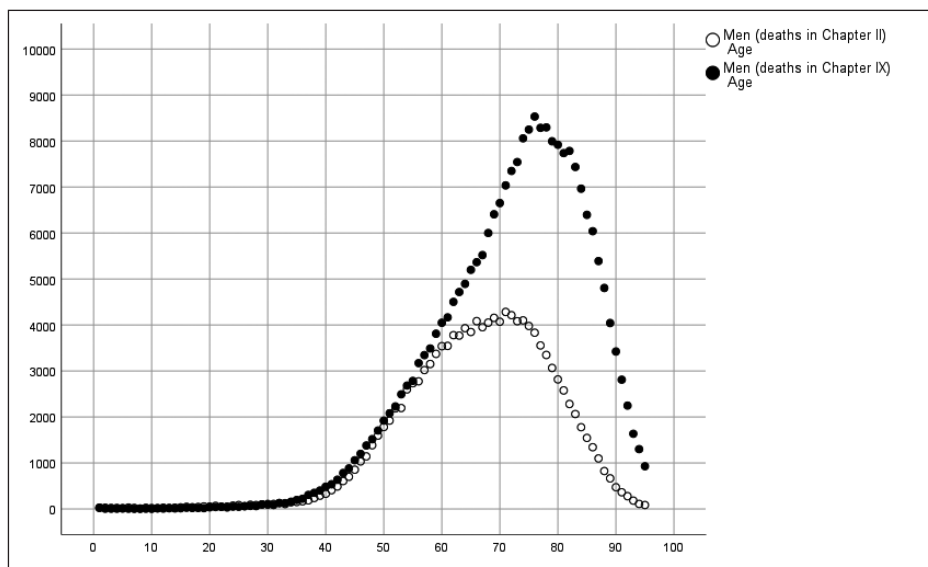
## RESULTS

### Curve Fitting of Mortality Counts

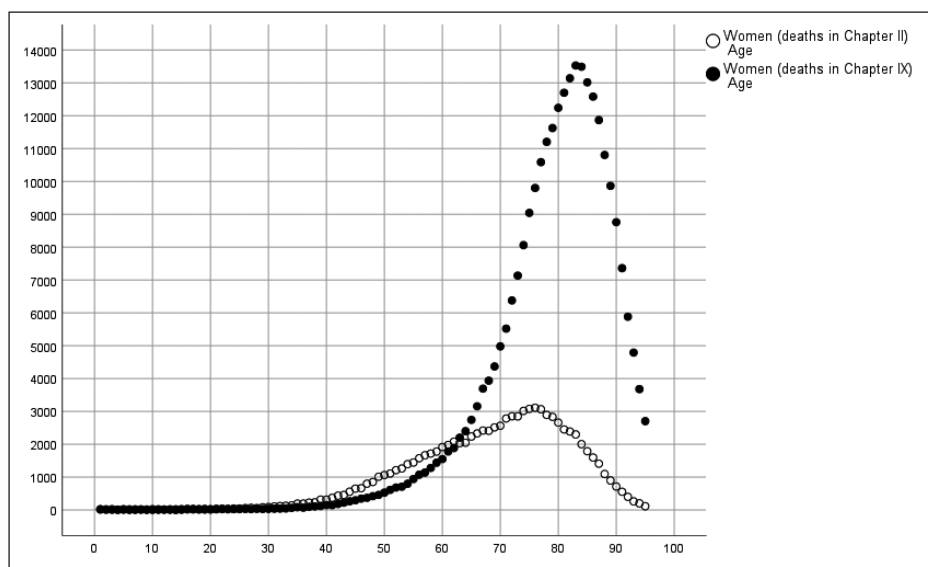
In Figure 1 are depicted absolute numbers of deaths caused by circulatory system diseases in comparison to numbers of deaths caused by neoplasms of Slovak men during 1996–2014 in one year age steps from 0 to 100 years. Cases above 100 years were excluded due to low occurrence. Numbers of deaths were absolute but were related to age. Therefore, influence of the age on absolute mortality counts was taken into account. The shape of both curves was rather smooth with exception of some irregular changes in cases of Chapter IX around age of 80 years. Number of deaths was smaller in Chapter IX in comparison to Chapter II up to 31 years with exception of several years. On the other hand after the age of 31 years the number of deaths caused by diseases of Chapter IX was larger for all following years. Up to the 60 years of age the shapes of both curves were almost the same. Afterwards, the number of cases from Chapter IX rapidly increased each year, up to the maximum absolute count ( $D=8,531$ ) in the age of 76 years. It was the mode of Chapter IX deaths. Subsequently, the number of Chapter IX deaths decreased; slowly at first, then after the 85 years mark much more rapidly. It could be expected, due to the number of men after 85 years being relatively small. Maximum absolute count of Chapter II deaths was in the age of 71 years ( $D=4,283$ ). Hence, there was 5-year shift in modes between Chapter IX and Chapter II. It suggested that maximum absolute count of deaths caused by the diseases of Chapter IX was five years later than in case of deaths caused by the diseases of Chapter II.

Concerning the female data, the absolute mortality was different in both Chapters. Figure 2 displayed absolute numbers of deaths caused by circulatory system diseases in comparison to numbers of deaths caused by neoplasms of Slovak women during 1996–2014 in age interval from 0 to 100 years. Cases above 100 years were again excluded. The shape of both curves was smoother than in case of men. Number of deaths was smaller in Chapter IX, in comparison with Chapter II up to 63 years, with only one exception of individuals under one year of the age (case of infant mortality). After the age of 62 years the number of deaths caused by diseases of Chapter IX vs Chapter II was larger for all years. It should be noted that a shift occurred at 31 years of age, identified as larger deaths counts caused by the diseases of Chapter IX vs Chapter II, as compared to men (intersections of curves for age of 63 years vs. 32 years). After the age of 62 years the number of cases from Chapter IX increased rapidly, to its maximum absolute number ( $D=13,529$ ) at the age of 83 years (mode of Chapter IX deaths). Regarding ages higher than 83 years there was in steep decline. The maximum absolute number of Chapter II deaths ( $D=3,110$ ) was in the age of 76 years. The shift in modes was 7 years.

With regards to the odds ratio of mortality cases compared in chapters  $D_{IX}/D_{II}$ , Figure 3 depicted the corresponding ratio for



**Fig.1.** Age dependence of absolute numbers of deaths caused by circulatory system diseases (Chapter IX) in comparison to numbers of deaths caused by neoplasms (Chapter II) of Slovak men during 1996–2014.



**Fig.2.** Age dependence of absolute number of deaths caused by circulatory system diseases (Chapter IX) in comparison to number of deaths caused by neoplasms (Chapter II) of Slovak women during 1996–2014.

men. Age interval was restricted to the age of 95 years. It was necessary because the numbers of living men in high age groups was too small for reliable estimation. It should be noted that course of a curve is rather smooth, with monotonous increase – with exception of the age interval from appx. 40 years to appx. 60 years of age. As it was already mentioned, the number of cases in Chapter IX in comparison with cases in Chapter II was higher from the age of 32 years. From this age the ratio Chapter IX/Chapter II was always larger in a group of men. At the age of 90 years the ratio was approximately ten. It suggested that the probability of death caused by circulatory system diseases was ten times larger compared to the probability of death caused by the neoplasms. Concerning the age involvement, several trend curves from IBM SPSS were tested and exponential model was the best one. Exponential model of dependence of death counts

ratio caused by Chapter IX divided by death counts caused by Chapter II on age in group of men was:

$$MenratioIX / II = e^{0.073 \cdot (Age - 63.1)} \quad (2)$$

Coefficient of determination was very good (90.2%). Estimated exponential curve was solid line. Odds ratio values oscillate around one, up to the age of appx. 50 years. A critical age of 55 years should be noted. Odds ratio values were significantly and permanently larger than one, for all ages larger than 55 years ( $p < 0.001$ ). Finally, odds ratio values increase up to values of appx. 12.

Figure 4 presented results for women. We see that sequence of values was different than it was in case of men. We can say that up to 40 years of age the ratio was constant. Critical age was 62

years. Up to 62 years the odds ratio was smaller than one. But at the age of 63 years, odds ratio was for the first time significantly larger than one ( $p=0.012$ ) and it was true for all following larger age values. Afterwards, it increased up to values of more than 20. Exponential model of dependence of ratio Chapter IX/Chapter II on age in group of women was:

$$\text{Womenratio IX / II} = e^{0.105 \cdot (\text{Age} - 65.8)} \quad (3)$$

Coefficient of determination was excellent (98.8%). Estimated exponential curve was again solid line.

Furthermore, we present values of absolute counts of deaths together with corresponding mortality odds ratios in higher age (see Table 1). The mortality odds monotonously grow with increasing age. The differences between sexes were also noteworthy. In case of men the ratio was larger than one already in the age of 60 years but that was not the case of women. In group of men the value of ratio slowly increased (from 1.14 at the age of 60 to 7.25 at the age of 90 years). Relative increase was 6.36 in men ( $7.25/1.14$ ). In group of women the ratio was smaller but increased more rapidly (from 0.81 at the age of 60 to 12.27 at the age of 90 years). Relative increase was 15.15 in women ( $12.27/0.81$ ).

### Decision Tree of Mortality Counts

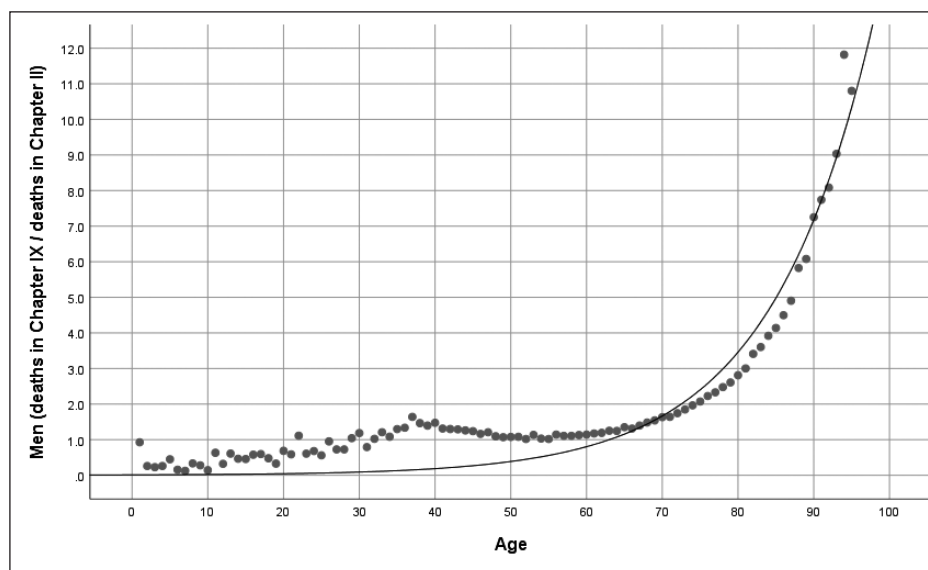
Our second research problem was to examine which available demographic determinants (age, region, sex, marital status and calendar year of death) significantly influenced the number of deaths caused by the diseases of the circulatory system in comparison with the number of deaths caused by neoplasms. We used IBM SPSS CART decision tree in default settings for the problem (Gini measure, five levels of maximum tree depth, etc....) with exception of equal prior probabilities, one standard error pruning and minimum parent (child) number of cases 1,000 (500) (see Fig. 5).

- At the beginning in node 0 the mortality odds of the diseases of the circulatory system (Chapter IX) vs diseases of neoplasms (Chapter II) was 2.35 (536,393/228,009).

**Table 1.** Absolute death cases caused by the circulatory system diseases ( $D_{IX}$ ) and by neoplasms ( $D_{II}$ ) together with their odds ratios (Ratio) for some ages

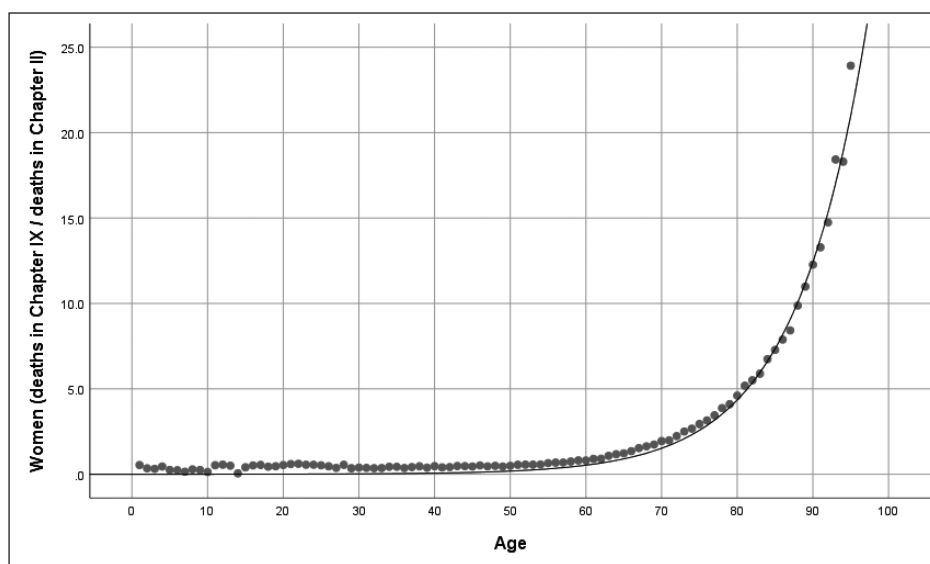
Age	Men			Women		
	$D_{II}$	$D_{IX}$	Ratio	$D_{II}$	$D_{IX}$	Ratio
60	3,539	4,048	1.14	1,916	1,549	0.81
65	3,848	5,199	1.35	2,236	2,742	1.23
70	4,071	6,651	1.63	2,562	4,980	1.94
75	3,979	8,249	2.07	3,075	9,042	2.94
80	2,817	7,919	2.81	2,658	12,240	4.60
85	1,545	6,395	4.14	1,788	13,016	7.28
90	472	3,423	7.25	714	8,760	12.27
95	86	929	10.80	113	2703	23.92

- At first level the first division was made by the age of death. Left node 1 contained 381,443 cases of age of no more than 75 years. The mortality odds were 1.35 (218,807/162,636). In the right terminal node 2 were cases with the age of at least 76 years and mortality odds of 4.86 (317,586/65,373), which was significantly higher than in the left node 1. It suggested that mortality odds of deaths caused by the diseases of the circulatory system in comparison with the number of deaths caused by neoplasms was larger for people of the age of at least 76 years, as compared to younger population. Additionally, the overall number of deaths was almost the same in both age groups. Hence, half of deaths caused either by diseases of Chapter II or by diseases of Chapter IX occurred before 76 years of age.
- The node 1 was again divided by the age at second level of the tree. If the age was higher than 66 years, then mortality odds were 1.02 (102,887/101,272 – node 3). Therefore, the probability of neoplasm-related death or circulatory-system-related death was the same in people under 67 years. If the age was higher than 66 years, then mortality odds were 1.89



**Fig. 3.** Age dependence of absolute number of deaths caused by circulatory system diseases (Chapter IX) divided by number of deaths caused by neoplasms (Chapter II) of Slovak men during 1996–2014.





**Fig. 4.** Age dependence of absolute number of deaths caused by circulatory system diseases (Chapter IX) divided by number of deaths caused by neoplasms (Chapter II) of Slovak women during 1996–2014.

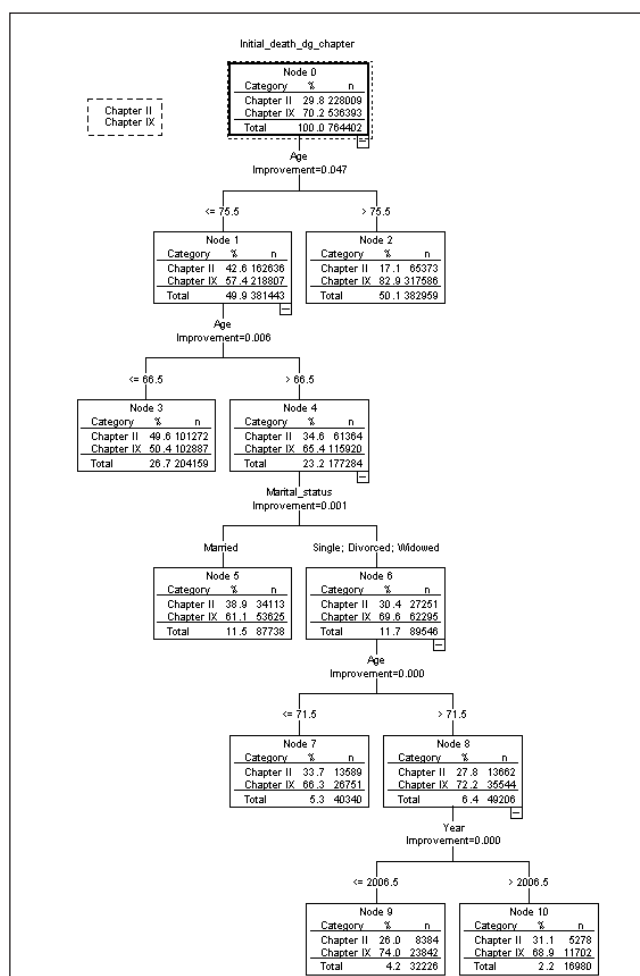
(115,920/61,364 – node 4). It was significantly higher than in the left node 3. However, the node 4 was concerned with people in age interval from 67 to 75 years.

- The next division was made by marital status at third level. Right node 4 was divided into cases from the group of married people with mortality odds of 1.57 (53,625/34,113 – left terminal node 5) and cases of people with other marital status with significantly higher mortality odds of 2.29 (62,295/27,251 – node 6). Mortality odds of deaths caused either by diseases of Chapter II or by diseases of Chapter IX, were significantly larger in the group of either single, divorced or widowed people in age interval from 67 to 75 years in comparison to married people. The overall number of deaths was almost the same in both groups. Hence, half of deaths caused either by diseases of Chapter II or by diseases of Chapter IX occurred in group of married people.
- At fourth level node 6 is further divided by age into cases with age of no more than 71 years (mortality odds = 1.97 (26,751/13,589 – terminal node 7, age interval from 67 to 71 years) and into cases with age of more than 71 years – mortality odds = 2.60 (35,544/13,662 – node 8, age interval from 72 to 75 years).
- The last division was made in node 8 by the calendar year at fifth level. In the left terminal node 9 are cases up to 2006 with mortality odds 2.84 (23,842/8,384, from 1996 to 2006) and in the right terminal node 10 are remaining cases after the year 2006 – mortality odds are 2.22 (11,702/5,278, from 2007 to 2014). However, this data represented only people in narrow age interval from 72 to 75 years.

The classification table of mortality odds of diseases of Chapter II in comparison with diseases of Chapter IX displayed that the CART decision tree model classified 64.8% of all cases correctly (see Table 2). The classification rate of the neoplasms cases was 67.7%. In the group of deaths caused by the diseases of the circulatory system it was 63.7%. We can see the prediction strength of the CART decision tree model. It was found that one need to ask only five questions to get correct classification of more than 63% of the mortality odds of diseases of the circulatory system vs neoplasms diseases.

Three questions were concerned with the age of death and two remaining with marital status and calendar year of death.

With regards to the relatively homogenous subgroups the following has been suggested:



**Fig. 5.** CART decision tree of mortality odds of diseases of the circulatory system vs. neoplasms.

**Table 2.** Classification matrix of mortality odds of mortality in diseases of the circulatory system (Chapter IX) vs neoplasms (Chapter II)

Observed	Predicted		
	Chapter II	Chapter IX	Correct (%)
Chapter II	154,252	73,757	67.7
Chapter IX	194,965	341,428	63.7
Overall (%)	45.7	54.3	64.8

Two subgroups on basis of only the age: people older than 75 years with mortality ratio 4.86 (node 2) and people younger than 67 years with mortality ratio 1.02 (node 3).

Two subgroups based on both age and marital status: married people in age interval from 67 to 75 years with mortality ratio 1.57 (node 5) and people in age interval from 67 to 71 years with mortality ratio 1.97 (node 7) with remaining marital status.

Two last subgroups were based on three predictors: age, marital status and calendar year of death. First two values were the same in both subgroups: age interval from 67 to 75 years and marital status single, divorced or widowed. They differed only in calendar year of death. The first subgroup contained older cases before 2006 (mortality ratio 2.84, node 9). The second one contains newer cases after 2006 (mortality ratio 2.22, node 10).

Importance (both relative and absolute) of all five predictors involved in CART decision tree model were depicted in Figure 6.

It could be suggested that the most important predictor in the mortality odds of the circulatory system diseases vs neoplasms diseases was the age of the death, followed by marital status (52.8%), followed by sex (14.2%). The last two predictors (year and region) were relatively negligible though formally significant.

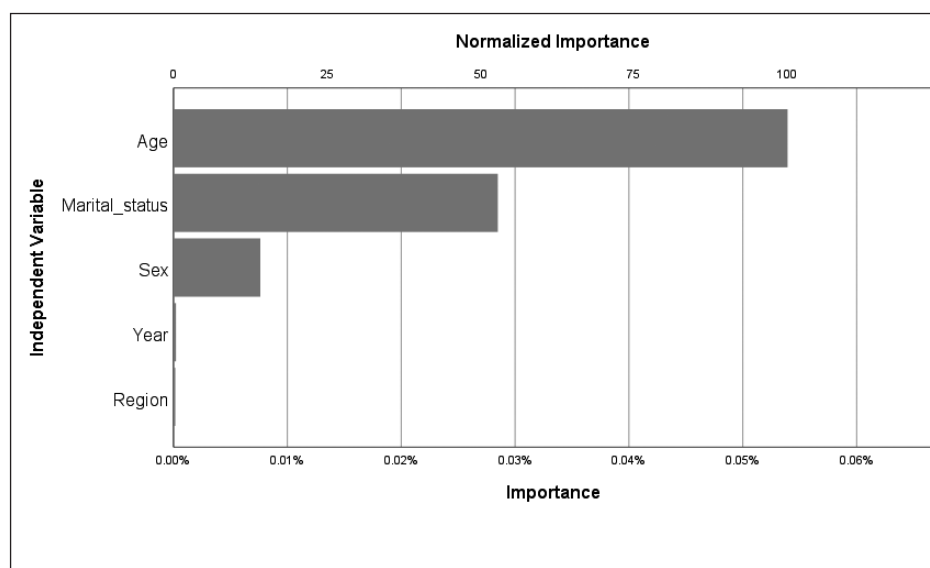
## DISCUSSION

Concerning sexes, the odds ratio in Chapter IX vs. Chapter II were almost constant, up to approximately 60 years, increasing

with higher age. We do not claim that it was true for all other possible odds ratios as well; however, concerning Chapter IX vs. Chapter II it should be obvious (see coefficients of determination). Furthermore, through involvement of age to the regression of odds ratio, the problem of age standardization could be solved as well. Regarding the decision tree, an important difference must be noted. In our analyses, we did not look at odds ratio as it is in the case of case-control studies. We used decision tree for classification between cases of Chapter IX and of cases of Chapter II.

Similar overall population studies were very important but we must also admit restrictions of their results due to other important determinants not being available. At an individual level, there were other factors (lifestyle, living conditions, human biology and genetics) that have not been considered. At macro level, we could mention population composition, medical care provision and environmental conditions (22–24). Despite these limitations, our results might be a valuable baseline for creating the predictive models.

Changes in the prevalence of risk factors of circulatory system diseases and neoplasms could be achieved by interventions on population and individual level. According to a large number of oncologists, Slovakia falls behind other European countries in cancer screening. Screening programs should have an organized form, preferably a full screening for well-defined target groups of people. However, it requires good organization and better financing. Unfortunately, there is no active screening program for cancer in Slovakia at this time. Similarly, for circulatory system diseases, the last prevention program “The national program for the prevention of diseases of the heart and blood vessels” has taken place in 2011 (25). This included a variety of educational activities, such as TV spots, leaflets, banners, as well as measuring cholesterol and blood pressure. The project was accompanied by the campaign “Know the symptoms of heart attack and stroke”, where health professionals explained symptoms and answered the questions from people in shopping centers. It is necessary to emphasize the importance of regular screening and prevention programs at both general and individual level, notably by identifying individuals at high risk.



**Fig. 6.** Plot of importance of demographic determinants of mortality odds of diseases of the circulatory system vs neoplasms.

## CONCLUSIONS

In our paper, we wanted to evaluate the influence of available demographic determinants to cases of death due to the circulatory system diseases (Chapter IX) in comparison to cases of neoplasms-related deaths (Chapter II). We used two approaches to achieve our objective. The first approach was trend curve fitting of mortality counts as function of the age for both sexes. We have found that at least in some cases it was possible to compute mortality odds ratio between two different diagnoses (or in our case groups of diagnoses). Afterwards, we could estimate mortality odds as a function of the age. This way we consider age as potential determinant and solve the problem of age standardization. In our study, we have found that ratio of deaths caused by the circulatory system diseases (Chapter IX) in comparison with deaths caused by neoplasms (Chapter II) were non-decreasing. At first for smaller age values the ratio of deaths was constant, but then it increased with increasing age (probable influence of other factors). In group of men the value of ratio was increasing from larger values, but it increased slowly (from 1.14 at the age of 60 to 7.25 at the age of 90 years). Relative increase was 6.36 in men (7.25/1.14). In group of women the ratio was smaller but it increased more rapidly (from 0.81 at the age of 60 to 12.27 at the age of 90 years). Relative increase was 15.15 in women (12.27/0.81). Therefore, relative risk of death caused by diseases of Chapter IX vs Chapter II was larger in group of women in higher age values.

The decision tree model was our second approach. Using this model, we have found that most significant demographic determinant of death counts in both ICD chapters was the age of deceased. The second one was marital status and the third one was sex. The last two predictors (year and region) were relatively negligible though formally significant.

## Acknowledgement

The work was supported by the VEGA Project No. 1/0993/15 "Genetic aspects in etiology of dilated cardiomyopathy".

## Conflict of Interest

None declared.

## REFERENCES

1. Ibrahim NA, Abdul Kudus A, Daud I, Abu Bakar MR. Decision tree for competing risks survival probability in breast cancer study. *IJMBR*. 2008;3(1):25-9.
2. Dudoit S, Fridlyand J, Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data. *J Am Stat Assoc*. 2002;97(457):77-87.
3. Loh WY, He X, Man MA. Regression tree approach to identifying subgroups with differential treatment effects. *Stat Med*. 2015;34(11):1818-33.
4. Jazbec J, Todorovski L, Jereb B. Classification tree analysis of second neoplasms in survivors of childhood cancer. *BMC Cancer*. 2007 Feb 2;7:27. doi.org/10.1186/1471-2407-7-27.
5. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, et al. A risk model for prediction of lung cancer. *J Natl Cancer Inst*. 2007;99(9):715-26.
6. Etzel CJ, Kachroo S, Liu M, D'Amelio A, Dong Q, Cote ML, et al. Development and validation of a lung cancer risk prediction model for African-Americans. *Cancer Prev Res*. 2008;1(4):255-65.
7. Tsien CL, Fraser HS, Long WJ, Kennedy RL. Using classification tree and logistic regression methods to diagnose myocardial infarction. *Stud Health Technol Inform*. 1998;52(1):493-7.
8. Solinger AB, Rothman SI. Risks of mortality associated with common laboratory tests: a novel, simple and meaningful way to set decision limits from data available in the electronic medical record. *Clin Chem Lab Med*. 2013;51(9):1803-13.
9. Pittman J, Huang E, Dressman H, Horng CF, Cheng SH, Tsou MH, et al. Integrated modeling of clinical and gene expression information for personalized prediction of disease outcomes. *Proc Natl Acad Sci USA*. 2004;101(22):8431-6.
10. Gülkesen KH, Köksal İT, Özden S, Saka O. Prediction of prostate cancer using decision tree algorithm. *Turk J Med Sci*. 2010;40(5):681-6.
11. Ignatius TS, Chiu YL, Wong TW, Tang JL. Deaths from nasopharyngeal cancer among waiters and waitresses in Chinese restaurants. *Int Arch Occup Environ Health*. 2004;77(7):499-504.
12. Clary T, Ritz B. Pancreatic cancer mortality and organochlorine pesticide exposure in California, 1989-1996. *Am J Ind Med*. 2003;43(3):306-13.
13. Lee LJ, Chung CW, Ma YC, Wang GS, Chen PC, Hwang YH, et al. Increased mortality odds ratio of male liver cancer in a community contaminated by chlorinated hydrocarbons in groundwater. *Occup Environ Med*. 2003;60(5):364-9.
14. MacArthur AC, Le ND, Abanto ZU, Gallagher RP. Occupational female breast and reproductive cancer mortality in British Columbia, Canada, 1950-94. *Occ Med*. 2007;57(4):246-53.
15. Pol L, Thomas R. The Demography of health and healthcare. Dordrecht: Springer; 2013.
16. Siegel KJ, Swanson DA. The Methods and materials of demography. Dordrecht: Springer; 2004.
17. Miettinen OS. Theoretical epidemiology: principles of occurrence research in medicine. New York: John Wiley & Sons; 1985.
18. Van Belle G, Fisher LD, Heagerty PJ, Lumley T. Biostatistics: a methodology for the health sciences. 2nd ed. Hoboken: John Wiley & Sons; 2004.
19. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. Oxford: Blackwell Science; 2002.
20. Breiman L. Classification and regression trees. New York: Chapman & Hall; 1984.
21. Šoltés V, Gavurová B. The possibilities of day surgery system development within the health policy in Slovakia. *Health Econ Rev*. 2014;4:35. doi:10.1186/s13561-014-0035-1.
22. Šoltés V, Gavurová B. The Functionality comparison of the health care systems by the analytical hierarchy process method. *E M Ekon Manag*. 2014;17(3):100-17.
23. Gavurová B, Vagašová T. Regional differences of standardised mortality rates for ischemic heart diseases in the Slovak Republic for the period 1996-2013 in the context of income inequality. *Health Econ Rev*. 2016;6:21. doi:10.1186/s13561-016-0099-1.
24. Kamenský G, Dvoranová I, Murín J, Dlesk A. The national program for the prevention of diseases of the heart and blood vessels [Internet]. 2011 [cited 2016 Nov 2]. Available from: <http://www.tvojesrdce.sk/files/documents/31.pdf>. (In Slovak.)

Received January 26, 2017

Accepted in revised form December 19, 2017