

Extension of the Results of a Prevention Guides for Cardiovascular Diseases Using Interpolation

Dorin Berian, Toma L. Dragomir

University “Politehnica” of Timișoara, Faculty of Automatics and Computers
Science, Bd. V. Pârvan 2, 300223 Timișoara, Romania, dorin.berian@aut.upt.ro,
toma.dragomir@aut.upt.ro

Abstract: For various diseases there exist databases which allow different correlations and estimations. Essentially, performing of such kind of operations means a systemic interpretation, some variables being considered to be cause variables, others being considered to be effect variables. The study is performed from this perspective over several data made public by the European Society of Cardiology in collaboration with other similar European societies towards the elaboration of prevention guides for cardiovascular diseases. It is developed an interpolative model which permits the risk estimation in more diversified circumstances than the initial information, thus the results being extended. It is discussed the model structuring and the interpolation method influence over its precision.

Keywords: cardiovascular diseases risk, interpolative type model, Matlab-Simulink implementation, cubic spline interpolation

1 Introduction

Medical data are recorded for a variety of purposes. They may be needed to support the proper care of the patient they were obtained from, but they also may contribute to the good of society through the aggregation and analysis of data regarding populations of individuals [1].

In order to perform studies regarding population's health status it is need to collect a large amount of medical data in databases. Obtaining such kind of databases implies long time and costly studies. These kind of studies aren't made public directly, but indirectly, through their results.

This paper uses the public results of a study that European Society of Cardiology in collaboration with other similar European societies performed in order to elaborate prevention guides for cardiovascular diseases (CVD) [2]. The result of the study mentioned above presents input data that have discreet character, thus it needs to use an interpolation method in order to apply them for any value of the input.

In this context, the paper has two main parts. Thus, considering the evaluation of the risk of cardiovascular disease to be a causal process, in section 2 the model of this process is built through the association of variables and linguistic values to the known data sets. The model type is interpolative, based on support points obtained from the primary data. Because the model is not finished until the proper interpolation method is chosen, in section 3 of the paper this problem is studied in order to implement the model in Matlab-Simulink.

2 Modeling of the Process for the Cardiovascular Diseases Occurrence Assessment

The starting point in this paper is based on the charts in Figure 1 [2] which refers to the evaluation of the ten year risk of fatal cardiovascular diseases occurrence for a patient living in one of the Europe regions with high risk for cardiovascular diseases occurrence. These data can be applied also for the Romanian patients because Romania is one of the Europe's countries with high risk for cardiovascular diseases occurrence [3].

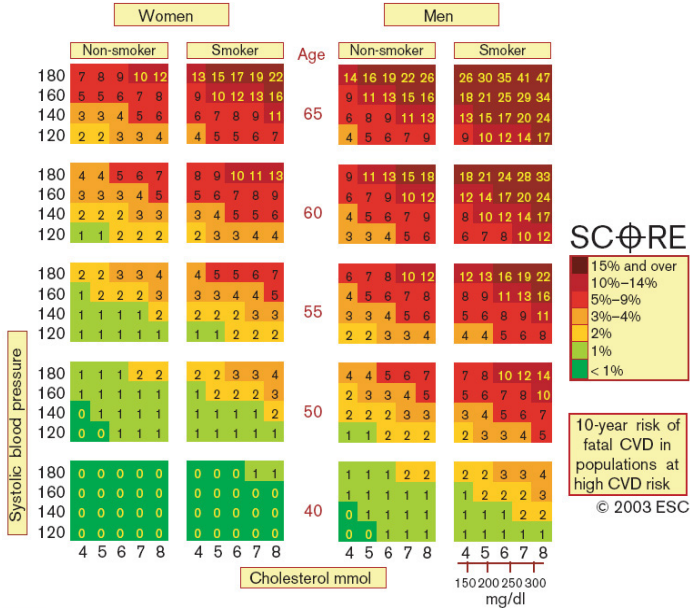


Figure 1

The ten-year risk of fatal cardiovascular diseases in high risk regions of Europe by gender, age, SBP, total cholesterol and smoking status

Based on these charts, the risk can be correlated causally with five attributes of the subject and can be approached systemically under this form:

$$\{Sex, Smoking\ status, Age, Systolic\ blood\ pressure, Cholesterol\} \rightarrow \{Risk\} \quad (1)$$

The information in the above charts defines a multivariable discrete system to the input. Globally, regarding the numerous patients, it can be considered that the information in the charts reflects a continuous process depending by the age, systolic blood pressure (SBP) and cholesterol. Taking into account that each information in the charts presented in Figure 1 can be assimilated with a support point in the six dimensional space of the input and output variables of the system (1) it was adopted the idea of an interpolative modelling of the system.

The five attributes of the subject represents inputs we consider linguistic values as in the table presented in Figure 2. In order to simplify the expression, certain linguistic values were named through alphanumerical symbols.

Linguistic variable	The name of the linguistic values
Sex	woman / man
Smoking status	non-smoker / smoker
Age	40 years / 50 years / 55 years / 60 years / 65 years
Systolic blood pressure (SBP)	120 mmHg / 140 mmHg / 160 mmHg / 180 mmHg
Cholesterol	4 mmol / 5 mmol / 6 mmol / 7 mmol / 8 mmol

Figure 2

Input variables and linguistic values associated with the system (1)

The risk level represents the output of the system and it is defined through percents related to the proper patient category. It varies between 0% and 47%, but using extrapolation, it can reach larger values. The risk level may be considered also as a linguistic variable, but the assimilation of the charts with a rules base that provides the support points don't imply this. In the following we will model the system in Matlab Simulink.

First, considering the bivalent structural character of the *sex* and the *smoking status* and the information related to the risk, the system (1) could be divided in four subsystems, each of them having as inputs the *age*, the *systolic blood pressure* and the *cholesterol* and as output the risk related to the considered structural category:

$$\{Age, SBP, Cholesterol\} \rightarrow \{Non-smoker\ women\ risk\} \quad (2.1)$$

$$\{Age, SBP, Cholesterol\} \rightarrow \{Smoker\ women\ risk\} \quad (2.2)$$

$$\{Age, SBP, Cholesterol\} \rightarrow \{Non-smoker\ men\ risk\} \quad (2.3)$$

$$\{Age, SBP, Cholesterol\} \rightarrow \{Smoker\ men\ risk\} \quad (2.4)$$

The associated Simulink model is presented in Figure 3. For the interpolation there are used four *Lookup Table 3D* blocks (one for each subsystem (2)) which operate with the support points from the charts in Figure 1 ($P_{s1s2ijk}$ represents a support point of the block corresponding to the *sex* $s1$ and the *smoking status* $s2$):

$$P_{s1,s2,ijk} = \{SBP_i, cholesterol_j, age_k, risk_{ijk}\} \quad (3)$$

The model provides the ten-year risk of fatal cardiovascular diseases for common values of the *systolic blood pressure*, the *cholesterol* and the *age*, limited to the values presented in Figure 2.

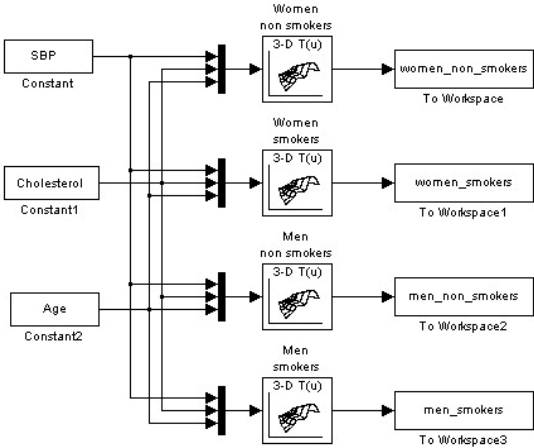


Figure 3

The Matlab Simulink model used for the evaluation of the ten-year risk of fatal cardiovascular diseases for a patient in one of the high-risk regions of Europe

Because the operating with charts having three inputs decreases the user's capability to interpret the correlations, the risk was represented depending by only two inputs, the third input being considered as a parameter. In Figure 4 it is represented the risk depending by the *systolic blood pressure* and the *age*, considering the *cholesterol* as a parameter (8 mmol). The parameter in Figure 5 is the *systolic blood pressure* (180 mmol), and the parameter in Figure 6 is the *age* (65 years). For a smoother graphical representation it was used the cubic spline interpolation. Obviously, the information can be structured in a different manner.

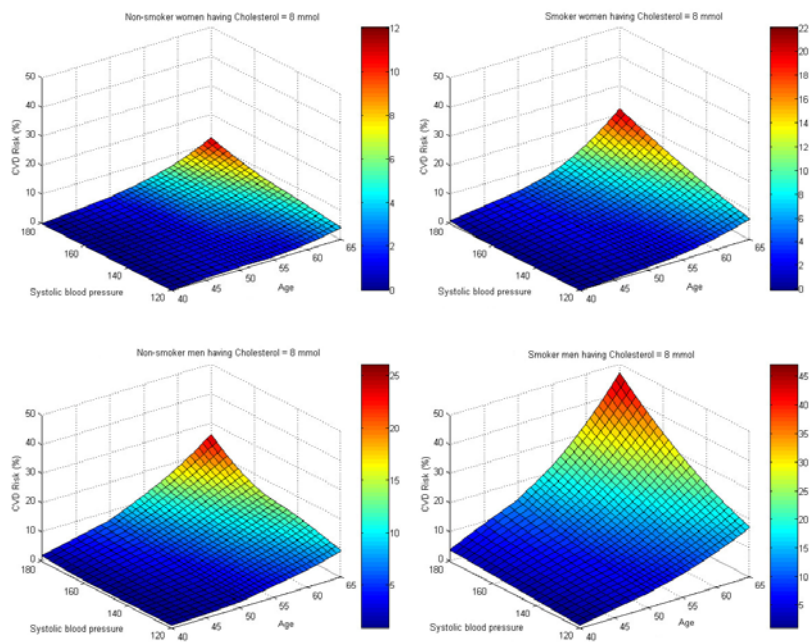
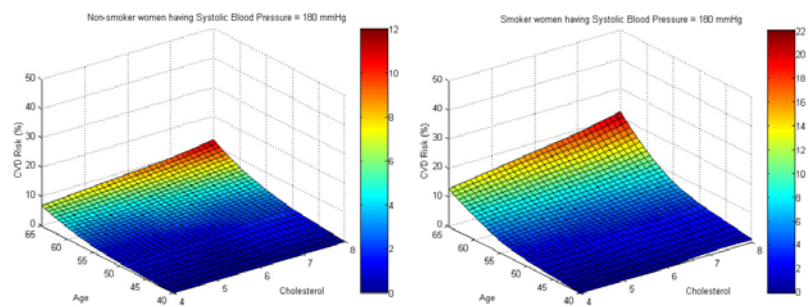


Figure 4
Risk variation with the SBP and age when the cholesterol = 8 mmol, for the following structural categories: non-smoker women, smoker women, non-smoker men and smoker men



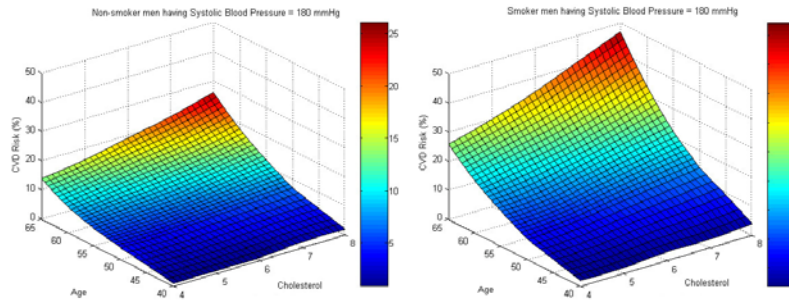


Figure 5

Risk variation with the age and the cholesterol when the SBP = 180 mmHg, for the following structural categories: non-smoker women, smoker women, non-smoker men and smoker men

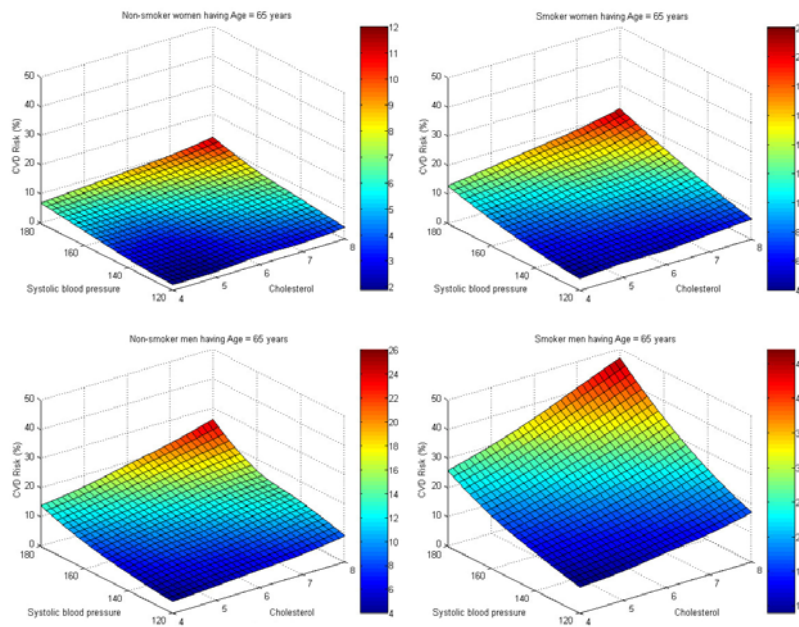


Figure 6

Risk variation with the SBP and the cholesterol when the age = 65 years, for the following structural categories: non-smoker women, smoker women, non-smoker men and smoker men

The classification of the smoking status in *smoker* and *non-smoker* is too simple. In reality there exist more categories, for example the *passive smokers* (those who don't smoke, but inhale smoke because they live between smokers), *occasional*

smokers, and *desperate smokers*. Accordingly, a comprehensive approach of the ten-year risk of fatal cardiovascular diseases estimation problem is to consider the smoking status as a continuous variable having values in a certain interval. That means that the values in the charts presented in Figure 1 are the margins of the interval, and for the intermediary statuses the results can be achieved using interpolation. Thus, we consider that the interval is [0%, 100%] and that it appears four new linguistic values, which insert in the table presented in Figure 2, the following change:

Smoking status	0% (non-smoker) / 25% (passive smoker) / 50% (occasional smoker) / 100% (desperate smoker)
----------------	--

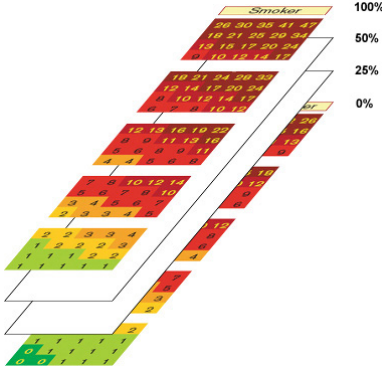


Figure 7

Considering the smoking status as a continuous variable having values in the interval [0%, 100%] we achieve charts with intermediary values of the values in the charts presented in Figure 1

In this context, the system can be restructured in order to have only two subsystems, each of them having four inputs:

$$\{Age, SBP, Cholesterol, Smoking\} \rightarrow \{Women Risk\} \tag{4.1}$$

$$\{Age, SBP, Cholesterol, Smoking\} \rightarrow \{Men Risk\} \tag{4.2}$$

In order to reflect this new situation, the Matlab Simulink model changes, the four *Lookup Table 3D* blocks being replaced by two *Lookup Tables 4D* blocks (as seen in Figure 8).

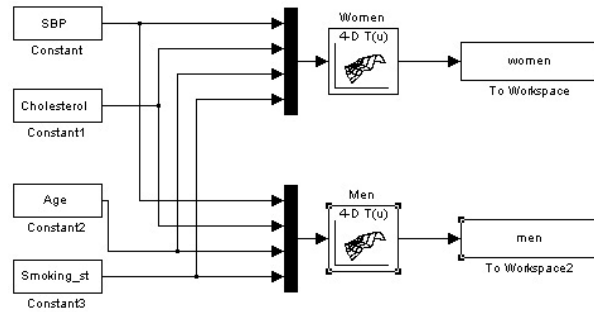


Figure 8

The new Matlab/Simulink model having *Lookup Table 4D* blocks used for the evaluation of the ten-year risk of fatal cardiovascular diseases

In this case, the support points the two blocks operate with are:

$$P_{x_l,ijkl} = \{SBP_i, Cholesterol_j, Age_k, Smoking\ status_l\} \quad (5)$$

3 The Influence of the Interpolation Method Over the Risk Evaluation

The values of the risk estimated through interpolation between the support points (3) and (5) using the models in Figures 3 and 8 are obviously dependent by the interpolation method used by the *Lookup Table 3D* or *Lookup Table 4D* blocks. In Simulink, the two interpolation methods available at the Lookup Table blocks are: cubic spline interpolation and linear interpolation. In this context, the question arises is the proper interpolation method that approximates the most correct values of the risk.

In order to approach more complex the problem of interpolation, beside the two interpolation methods, above mentioned it was also considered the Shepard interpolation (that uses weighting factors proportional with μ - power of the inverse distance between the current point and the support points). [4].

In order to identify the most accurate interpolation method, there were accomplished four tests for the estimation of the risk. In the frame of the each test, for the comparison of the three interpolation methods: *cubic spline*, *linear* and *Shepard* there were used the same support points resulted from the charts presented in Figure 1 through the omission of a set of support points chosen for the validation. In the four tests, for the Shepard interpolation, the μ parameter took values between 1 and 5.

For the assessment of the accuracy of the estimations for the risk r using the three interpolation methods there were compared the estimation errors relative to the reference values $\mathcal{E}_r = (r_{estimated_through_interpolation} / r_{charts_in_Fig.1}) - 1$.

In this context there were also compared other values: i) the arithmetic mean of the absolute values of the relative errors; ii) the maximal positive and negative values of the relative errors.

- **The first test** (several results related to the non-smoker patients are represented in Figure 9).

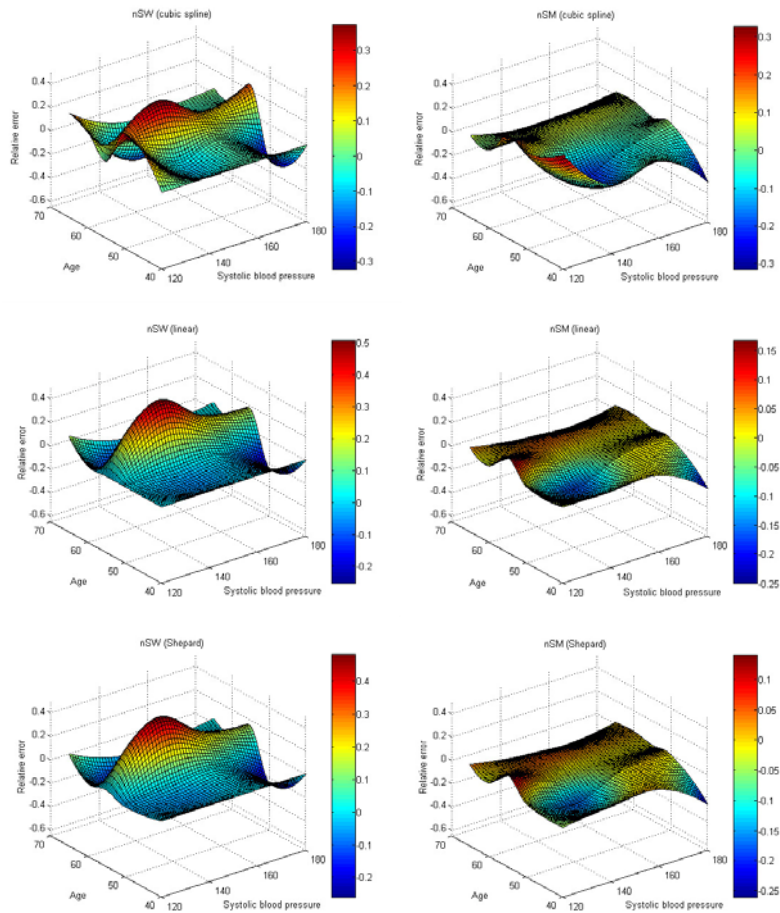


Figure 9

The relative errors achieved in the first test for the non-smoker women (the left column) and for the non-smoker men (the right column) for the three interpolation methods

Conditions: From the charts presented in Figure 1 we have eliminated all the support points corresponding to the *cholesterol of 7 mmol*.

Results: The most accurate estimations (the smallest relative errors) were achieved for the Shepard interpolation ($\mu= 3.5$) and the linear interpolation. For every interpolation method the maximal relative errors were achieved for small values of the risk (risk = {0, 1}), thus, taking into account their absolute values associated with a low risk level, these errors can be considered negligible.

▪ **The second test** (several results are represented in Figure 10)

Conditions: From the charts presented in Figure 1 we have eliminated all support points corresponding to the *age of 60 years* and the *cholesterol of 6 mmol*.

Results: The most accurate estimations (the smallest relative errors) were achieved for the cubic spline interpolation. For the Shepard interpolation the smallest relative errors of the risk were achieved for $\mu= 2.5$. For every interpolation method the maximal relative errors were achieved for reasonably small values of the risk (risk = {2, 3}), thus, as in the first test, these errors can be considered negligible.

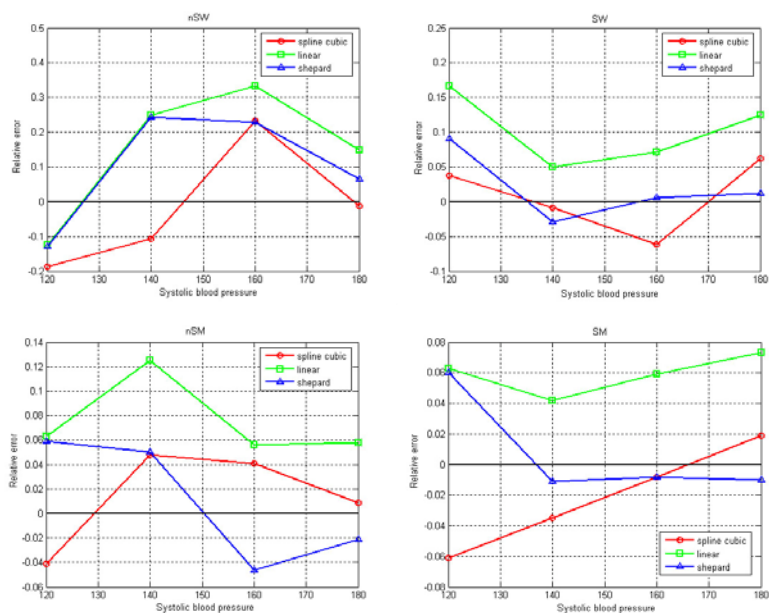


Figure 10

The relative errors achieved in the second test for each of the four categories (*non-smoker women, smoker women, non-smoker men and smoker men*). Each of the four graphics represents the results achieved for the three interpolation methods

- **The third test** (several results are represented in Figure 11)

Conditions: From the charts presented in Figure 1 we have eliminated all the support points corresponding to the *age of 55 years*, the *systolic blood pressure of 160 mmHg* and the *cholesterol of 6 mmol*.

Results: The smallest relative errors were achieved for the cubic spline interpolation. For the Shepard interpolation the smallest errors were achieved for $\mu=1$.

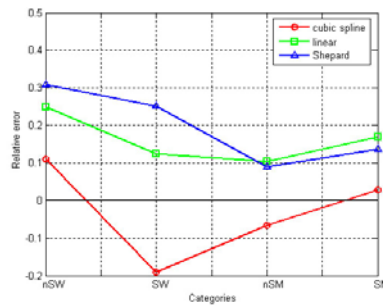
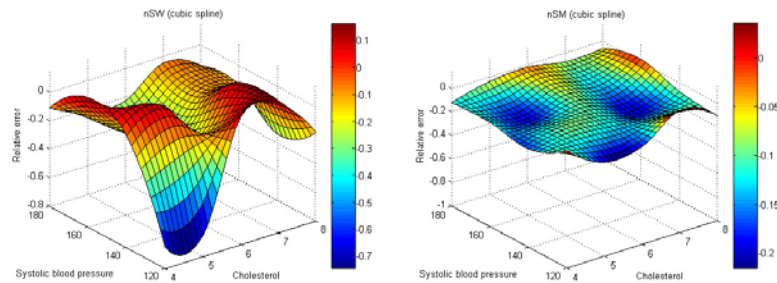


Figure 11

The relative errors achieved in the third test for all the four categories (*non-smoker women, smoker women, non-smoker men and smoker men*) for the three interpolation methods

- **The fourth test** (several results are represented in Figure 12)

Conditions: From the charts presented in Figure 1 we have eliminated the support points corresponding to the *age of 65 years*. In this case, unlike the precedent tests, there were eliminated from the charts boundary support points (associated with a limit of the interval of one of the inputs), thus considering that the risk will be estimated through extrapolation.



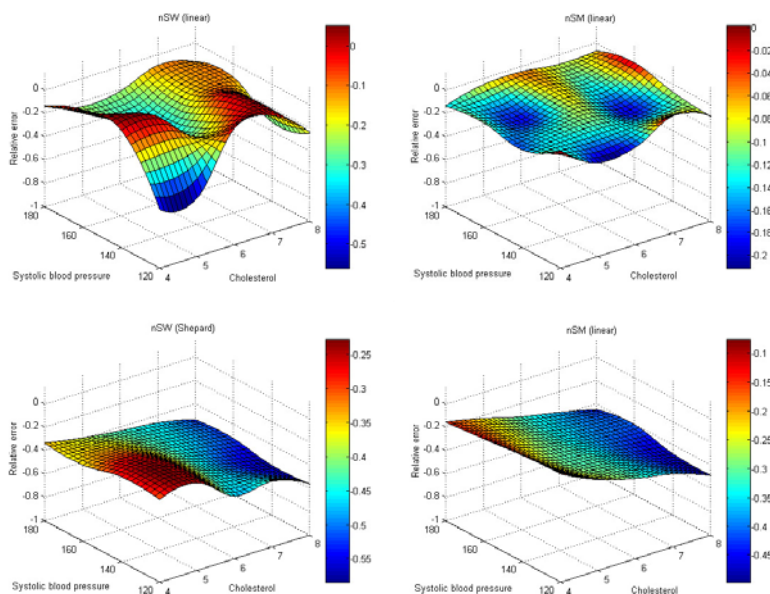


Figure 12

The relative errors achieved in the fourth test for the non-smoker women (the left column) and for the non-smoker men (the right column) for the three interpolation methods

Results: The smallest relative errors were achieved for the cubic spline interpolation. For the Shepard interpolation, the smallest errors were achieved for $\mu=5$.

Analyzing the results achieved for the three interpolation methods used in the four precedent tests, there can be stated the following conclusions:

The cubic spline interpolation: It is relevant that in three of the four tests, the best results were achieved using the cubic spline interpolation method. It is also important that the cubic spline interpolation is the only one which had produced relative errors equidistant distributed to zero. Consequently it can be stated that the spline cubic interpolation has the advantage to provide generally acceptable estimations for nonlinear variations of the risk.

The linear interpolation: In no one of the four tests, the linear interpolation provided the best estimations. However, in three of the four tests, the estimations achieved through linear interpolation were next to the best estimations (in the first test, the best estimations were achieved through Shepard interpolation, and in the third and fourth test the best estimations were achieved through cubic spline interpolation). Even if the estimations achieved through the linear interpolation don't have significant errors relative to the reference values of the risk, the linear

interpolation is not adequate for the heavy nonlinear variations of the risk (border domains).

The Shepard Interpolation: The results achieved through the Shepard interpolation are essentially dependent by the μ parameter. In order to obtain minimal relative errors, the μ parameter should be adjusted for each situation apart. So, it could not be used a unique value for μ . There could not be found any logical way to choose the optimal value for μ , other than through trials. In conclusion, this interpolation method cannot be generalized for any input.

Considering that in the fourth test the largest relative errors were related to the high levels of the risk, it can be stated that the Shepard interpolation may produce big estimation errors, even if in the first test, the estimations achieved through Shepard interpolation were the nearest of the reference values of the risk (those presented in the charts from Figure 1).

Also, the most proper interpolation method for the estimation of the ten-year risk of fatal cardiovascular diseases is the cubic spline interpolation.

Conclusions

The evaluation of the ten year risk of fatal cardiovascular disease for a population can be approached in a systemic mode, as a causal process having as inputs the following variables: *Sex*, *Smoking status*, *Age*, *Systolic blood pressure* and *Cholesterol* and as outputs the *Risk*. The model associated to the process is interpolative, tabular implemented through a set of support points. In order to use it easier it is divided in subsystems, each subsystem having two inputs, dependent by one or more parameters. The precision of the model is dependent by the interpolation method used, the most proper being proved to be the cubic spline interpolation.

References

- [1] Shortlife, E. H., Barnett, G. O., "Medical Data: Their Acquisition, Storage, and Use", Medical Informatics – Computer Applications in Health Care and Biomedicine, second edition, Springer NY 2001, pp. 49
- [2] European Society of Cardiology,
http://www.escardio.org/NR/rdonlyres/A0EF5CA5-421B-45EF-A65C-19B9EC411261/0/CVD_Prevention_03_full.pdf
- [3] European Heart Network,
www.ehnheart.org/files/EurCVDstat2000-112408A.pdf
- [4] Drechsel, D., "Regelbasierte Interpolation und fuzzy Control", Vieweg, 1996