

## Assessment of the Importance of a New Risk Factor in Prediction Models

Mohammad Reza Baneshi,<sup>1</sup> Ehsan Mosa Farkhani,<sup>2</sup> and Saiedeh Haji-Maghsoudi<sup>3,\*</sup>

<sup>1</sup>Research Center for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, IR Iran

<sup>2</sup>Department of Epidemiology, University of Tehran, Tehran, IR Iran

<sup>3</sup>Department of Biostatistics & Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, IR Iran

\*Corresponding Author: Saiedeh Haji-Maghsoudi, Department of Biostatistics & Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, IR Iran. Tel: +98-9133874358, Fax: +098-3413205405, E-mail: sa.maghsoudi@gmail.com

Received 2014 June 15; Revised 2014 September 3; Accepted 2014 September 28.

### Abstract

**Background:** Discovery of new risk factors poses new challenges on how to quantify their added value and importance in risk prediction improvement.

**Objectives:** The aim of this study was to apply different statistics and to quantify the importance of some risk factors in acute myocardial infarction (AMI).

**Patients and Methods:** In a retrospective cohort study, 607 patients with AMI, aged more than 25 years were studied. They were admitted to the CCU of Imam Reza hospital in Mashhad, Iran from 2007 to 2012. Health information and death registration systems were used to identify patients and to assess their outcome. At first a model containing all variables was fitted (full model). Importance of variables was compared in terms of standardized regression coefficient and inclusion frequency in bootstrap samples. Then, a series of reduced models were fitted, where in each of them only one of the independent variables was excluded. Models were compared in terms of goodness of fit, accuracy (Cindex, R square), separation of patients into risk groups (SEP), and net reclassification improvement (NRI).

**Results:** Age was selected as the important factor based on all 7 statistics. Exclusion of age variable decreased C index from 0.75 to 0.68 and R square from 0.25 to 0.15. Duration of hospitalization was important based on 4 statistics. Exclusion of this variable decreased R square from 0.25 to 0.21. While gender was a useful variable in separation of patients into risk groups, its omission did not reduce model likelihood. The opposite was true in the case of using streptokinase during hospitalization.

**Conclusions:** Our results revealed that a variable with high separation ability might not necessarily be useful in terms of goodness of fit. Therefore, importance should be defined carefully based on clinical objectives of the study.

**Keywords:** Acute Myocardial Infarction, Variable Importance, Added Value, Net Reclassification Improvement, Cindex, Survival Analysis

## 1. Background

Many studies have been funded to identify important prognostic factors for a disease course. Besides its financial burden, results of these trials are of value to execute new national health policies. The ongoing discovery of new risk factors poses new questions on how to quantify their contribution in risk prediction improvement (1). In other words, not only the significant association between new risk factor and event of interest but also its contribution in improvement of model performance should be addressed. A new significant risk factor is valuable when it improves the model performance (2, 3).

Different statistics are available to address the importance of a new risk factor. Importance can be defined in terms of magnitude of regression coefficient (standardized regression coefficient), stability of significance of risk factor against variations in the data (by assessing inclusion frequency of variables across bootstrap samples), goodness of fit of models with and without the risk fac-

tor of interest, predictive ability of models in the presence and absence of new risk factor (Harrell's Cindex, Nagelkerke Rsquare), separation of risk groups (SEP), or reclassification of patients into risk groups (net reclassification index (NRI)) based on comparison of classifications made by two models (2, 4-6).

These statistics quantify the importance of a variable based on different characteristics. Traditional method is to compare coefficient of risk factors. However, scale of measurement directly affects estimation of regression coefficients. Therefore, comparison of standardized regression coefficients are suggested (6).

Some authors argue that importance of a variable should be based on its stability against variations in the data as a weak predictor might lose its significance by inclusion or exclusion of a few subjects. Therefore, they suggest monitoring the significance of variables in data sets with minor differences (i.e. bootstrap samples). A

variable that remains significant in majority of bootstrap samples is stable and can be considered as an important risk factor (7).

Harrell's C index and Nagelkerke R square quantifies the predictive ability of a model. Thus, comparison of these statistics for models with and without a new risk factor provides a measure of its predictive ability. Comparison of likelihoods of model is also another measure of goodness of fit. On the other hand, recent developments suggest the idea of reclassification of patients across risk groups by different models (1-3, 8).

## 2. Objectives

Each of these approaches has its own weaknesses and strengths. The aim of this study was to introduce these statistics, and to explore whether an important risk factor from goodness of fit prospect, say, would be selected as being important from another prospect. To do so, we used a 5-year follow-up data of patients with acute myocardial infarction (AMI).

## 3. Patients and Methods

Imam Reza hospital in the heart of Mashhad metropolitan, Iran is a governmental referral hospital with 856 active beds and 18 wards. A total of 607 patients diagnosed with AMI from 2007 to 2012 period were followed by their hospital records. The event of interest was death. Time was defined as the difference between date of diagnosis and date of death (for died patients), or last follow-up time recorded in the patients' record (for censored subjects).

### 3.1. Case Ascertainment and Diagnostic Criteria

Hospitalized cases were identified from hospital medical records and general practitioners' reports. Validations of events were based on medical history, symptoms, ECG, and cardiac enzymes, and in some cases, general practitioners' reports. Patients were classified, according to the coding on the international classification of disease, 10th revision, (ICD-10 = I21.0 - I21.9). The validity of the diagnosis coding has a positive predictive value of 96% (9). Several exclusion criteria were applied to ensure the accuracy of the AMI diagnosis such as patients not admitted to CCU (0.9%), duplicate admission and intra-hospital transfers (1.4%), for patients transferred from another hospital, only the first admission counted (6.8%), and patients aged > 105 years (0%).

### 3.2. Mortality Follow-up

Vital statistics (i.e. date of death) was obtained by linking hospital admission data to the death registration system. Individuals not identified as deaths in the death registration system were assumed to be alive at the end of year 2012. Time was defined as the difference between date of diagnosis and date of death (for died patients), or last follow up time recorded in the patients' record (for censored subjects).

Independent variables were duration of hospitalization, gender, age, history of hypertension (yes, no), history of diabetes (yes, no), history of hyperlipidemia (yes, no), smoking status, history of ischemic heart disease (yes, no), family history (yes, no), addiction at diagnostic time (yes, no), Q wave in ECG (yes, no), and use of streptokinase during hospitalization (yes, no). All information was abstracted from hospital medical records and one observer collected them. These data were collected for another study (10) and we used them to illustrate our aim.

At the first step, a multifactorial Cox model was developed using all 12 independent variables (called full model). Importance of variables was explored using standardized regression coefficients, and inclusion frequency in bootstrap samples. Then, a series of reduced Cox models were fitted; in each of them, only one of the independent variables was omitted. Importance of variables was assessed comparing full and reduced models in terms of goodness of fit, model accuracy, separation ability, and reclassification of patients into risk groups. Details are provided below.

### 3.3. Standardized Coefficient

Standardized version of the variables was offered to the multifactorial model and regression coefficients were compared. Coefficients of multifactorial model were obtained from maximum likelihood method. The higher the coefficient, the more important the variable is.

### 3.4. Bootstrap Stability

A total of 100 bootstrap samples were drawn from the original sample. A full model was fitted to each of them. We then counted significance of each variable across 100 bootstrap samples. Clearly, variable with the highest inclusion frequency is less sensitive to the fluctuations in the data and can be considered as the most important predictor.

### 3.5. Goodness of Fit

Log likelihood is one of the measures used for determination of goodness of fit model. This statistics use the maximum likelihood function.

We calculated the difference between  $-2 \times \log$ -likelihood for the full and that of reduced models. This statistics follow a Chi-square statistics. We explored whether omission of any of independent variables significantly reduce the model likelihood.

equation of likelihood ratio test (LRT) is as follows:

$$(1) \quad LRT = -2 \log \frac{L(\text{MLE of restricted model})}{L(\text{MLE of unrestrained model})}$$

### 3.6. Accuracy (Discrimination and Predictive Ability)

Accuracy is the degree to which predictions match

outcomes (11). Accuracy was measured in terms of discrimination (Harrell's C index) and predictive ability (R square). In risk stratification studies, it is important to create risk groups where patients in each group are equally likely to develop the outcome (11). Discrimination refers to the ability to separate patients with different responses. Discrimination is measured using Harrell's C index (concordance index) which is a generalization of area under curve (AUC). C index is related to rank correlation between predicted and observed outcome and defined as the proportion of all usable patient pairs in which the prediction and outcome are concordant (12, 13). This statistic varies between 0.5 and 1 where values near 1 indicate high discrimination power (14).

On the other hand, Nagelkerke R square was used to compare the predictive ability of models. This statistics, which vary between 0 and 1, indicates the ability of a set of prognostic factors to predict outcomes. Values of 0 and 1 indicate very poor and very high predictive ability, respectively (15). We calculated Cindex and Nagelkerke Rsquare of full model. Then, we explored the impact of omission of each independent variable on these statistics.

### 3.7. Separation Ability

Multiplying the regression coefficients into variables, a risk score was calculated for each patient. Applying percentiles 33 and 66, three risk groups were created (low, intermediate, and high). Difference between actuarial survival rate between low and high risk groups at fifth year was defined as separation ability (SEP). This statistics was calculated for full and all reduced models.

### 3.8. Net Reclassification Index

A recent paper suggested the use of net reclassification index (NRI) as a tool to assess usefulness of a new risk factor. Suppose that full and reduced model classify patients to low, intermediate, and high risk groups. For patients who experienced the event, shift to higher risk groups means improvement. The opposite is true for those who did not experience the event. If D denotes event indicator, NRI is defined as below ( ≠ indicates number of):

$$(2) \quad P(up|D = 1) = \hat{P}_{up, events} = \frac{\neq events \text{ moving up}}{\neq events}$$

$$(3) \quad P(down | D = 1) = \hat{P}_{down, events} = \frac{\neq events \text{ moving down}}{\neq events}$$

$$(4) \quad P(up|D = 0) = \hat{P}_{up, non - events} = \frac{\neq non - events \text{ moving up}}{\neq non - events}$$

$$(5) \quad P(down | D=0) = \hat{P}_{down, non - events} = \frac{\neq non - events \text{ moving down}}{\neq non - events}$$

$$NRI = [P (up | D = 1) - P (down | D = 1)] - [P (up | D = 0) - P (down | D = 0)].$$

## 4. Results

Out of 607 patients enrolled in this study, 204 deaths occurred, giving death rate of 34%. The mean ± SD age of died and survived patients were 69.1 ± 11.7 and 57.4 ± 12.5 years, respectively.

Among died and survived patients, 55% and 33% had history of hypertension, respectively. Prevalence of diabetic among these groups was 37% and 18%, respectively. Corresponding figures for hyperlipidemia history were 28% and 19%. Prevalence of history of ischemic heart diseases in died patients was higher than survivors (41% vs. 25%, respectively).

In terms of inclusion frequency of variables in 100 bootstrap samples, variables reached significant levels in majority of samples were age (100), duration of hospitalization (100), and history of diabetes (93). The gender variable remained significant in slightly higher than half of samples (56) (Table 1).

Three of these variables had high standardized coefficient (age, duration of hospitalization, and history of diabetes). The fourth most important variable was use of streptokinase during hospitalization period.

While history of ischemic heart disease was not important in terms of inclusion frequency or standardized coefficient, deletion of this variable significantly would reduce goodness of fit of the full model. From this prospect, other important variables were age, duration of hospitalization, and history of diabetes.

Predictive ability of the full model was 0.25. We observed that exclusion of age dramatically changed model R square, where 10 percentage point difference was observed (0.15 versus 0.25). Exclusion of duration of hospitalization reduced predictive ability of the full model to 0.21. No remarkable reduction observed by deletion of any other variable (Table 1).

In terms of discrimination ability, C index for full model was 0.748. Exclusion of no variable made considerable change in C, except age that gave C index of 0.68. Exclusion of other variables only changed third decimal figures.

Difference between 5-year survival rate of low and high risk groups (SEP) created based on the full model was 0.54. Impact of age omission was considerable (SEP = 0.39). However, deletion of duration of hospitalization did not remarkably affect the separation ability of the model (SEP = 0.51). Gender variable, which was not important from other prospects, was essential in separation of risk groups (SEP = 0.49) (Table 1).

In terms of reclassification of patients into risk groups, only omission of age resulted in more than 10% misclassification rate (-14% for died and -6% for censored patients) (Table 1).

**Table 1.** Comparison of Importance of Independent Variables From Different Prospects<sup>a,b</sup>

Variable	Inclusion Frequency	Standardized Beta <sup>c</sup>	Difference in LR	R <sup>2</sup>	C Index	SEP	NRI, %	
							Censored	Died
All	-	-	-	0.25	0.748	0.54	-	-
Duration of hospitalization	100 <sup>d</sup>	-0.48 <sup>d</sup>	30.93 <sup>d</sup>	0.21 <sup>d</sup>	0.745	0.51	-1.7	-4
Gender	56	0.14	3.78	0.25	0.745	0.49 <sup>d</sup>	-2.0	-4.4
Age	100 <sup>d</sup>	0.78 <sup>d</sup>	87.05 <sup>d</sup>	0.15 <sup>d</sup>	0.680 <sup>d</sup>	0.39 <sup>d</sup>	-6.0 <sup>d</sup>	-13.7 <sup>d</sup>
History of hypertension	21	-0.08	1.23	0.25	0.747	0.51	-1.2	-2.5
History of diabetes	93 <sup>d</sup>	-0.24 <sup>d</sup>	12.06 <sup>d</sup>	0.24	0.740	0.52	-1.5	-2.9
History of hyperlipidemia	11	0.04	0.34	0.25	0.748	0.53	-0.3	-0.5
Smoking status	10	-0.01	0.51	0.25	0.746	0.50	-1.5	-2.9
History of Ischemic heart disease	33	-0.12	4.7 <sup>d</sup>	0.25	0.746	0.52	-1.0	-2.0
Family history	2	-0.01	0.02	0.25	0.748	0.54	0.3	0.0
Addiction	14	-0.05	0.6	0.25	0.748	0.52	-0.7	-1.0
Q wave in ECG	14	0.06	1.31	0.25	0.748	0.52	-0.7	-1.5
Use of streptokinase	38	0.16	0.36	0.25	0.746	0.51	-1.0	-2.0

<sup>a</sup>Abbreviations: LR, likelihood ratio; NRI, net reclassification index; SEP, separation ability.

<sup>b</sup>Negative values show percent misclassification in risk groups.

<sup>c</sup>Standardized beta shows coefficients of variables in the full model.

<sup>d</sup>Shaded cell shows important variables based on different prospects.

## 5. Discussion

Assessment of the importance of a new risk factor can be addressed using different statistics. We should emphasize that our goal was not to identify important risk factors of AMI. In this study, we compared application of 7 criteria in the content of survival analysis. All 7 criteria selected age as being an important risk factor for death of the patients with AMI. Duration of hospitalization and history of diabetes were important based on 4 and 3 statistics, respectively. On the other hand, gender and history of ischemic heart disease was significant only from 1 prospect.

C statistics is the most frequently statistics used to quantify the added value of a new risk factor (13). Cook (16) highlighted that the impact of a new predictor on the C statistic is lower when other strong predictors are in the model, even when it is uncorrelated with the other predictors. Moreover, strong independent association of a new marker with outcome is required to result in meaningfully larger AUC (17-19). As an example, by adding a new biomarker to a set of standard risk factors predicting CVD, the model AUC would increase from 0.76 to 0.77 (20).

To evaluate the added value of HDL cholesterol in risk prediction of coronary heart disease, Pencina (3) added it to a standard prediction model. Addition of this variable improved the goodness of fit and reclassification of patients into risk groups. The AIC for old and new models were 2779 to 2762. The NRI was estimated at 0.12. The cor-

responding AUCs were 0.762 and 0.774, which were not statistically significant (3).

Uno et al. (21) assessed the importance of a new biomarker, “wound-response gene expression signature”, analyzing a breast cancer data set. C statistics with and without the gene score were 0.71 and 0.69, respectively. The improvement was 0.02 with 95% confidence interval (-0.01, 0.05), which was not significant. Authors did not calculate any other statistics.

Meigs et al. (22) showed that increase in C is much more obvious when C corresponded to the standard model is low rather than high. They calculated a genotyped score using single nucleotide polymorphisms (SNPs) at 18 loci associated with diabetes. In a model adjusted for sex and self-reported family history of diabetes, the C statistic for models with and without genotyped score was 0.595 and 0.615. On the other hand, in a model adjusted for age, sex, family history, body-mass index, fasting glucose level, systolic blood pressure, high-density lipoprotein cholesterol level, and triglyceride level, the C statistic was 0.900 without the genotype score and 0.901 with that score (P = 0.49).

Wang et al. (20) measured 10 biomarkers in 3209 participants attending a routine examination cycle of the Framingham heart study. The main outcome of the study was death. C statistics were 0.79 for predictors of age, sex,

and multi-marker score as, 0.80 for predictors of with age, sex, and conventional risk factors, and 0.82 with all predictors. As shown, inclusion of a score created by contribution of 10 biomarkers, only led to 3 percentage point increase in C statistics.

In the context of survival data, researchers wish to create risk groups as diverse as possible. We have seen that deletion of history of ischemic heart disease significantly reduced LRT, but did not affect SEP. The opposite was true in the case of gender variable.

We should emphasize that the criteria used to ascertain importance of variables should be selected based on clinical goal of the study. For example, AUC focuses only on the predictive accuracy of a model. As such, it cannot tell us whether the model is worth using at all. For example, suppose a scenario in which false-negative result is much more harmful than a false-positive result. A model with much greater specificity but slightly lower sensitivity than another model results in higher AUC but might not be preferred from clinical practice.

Among these statistics, in recent years, lots of attention has been paid to NRI. We believe that this statistics has some drawbacks. For instance, in management of breast cancer patients, Nottingham prognostic index is the gold standard. This index classifies cases into 3 groups as low, intermediate, and high risk. However, addition of a new biomarker might give the flexibility to divide intermediate risk groups into two separate sets as higher-intermediate and lower-intermediate risk functions. Therefore, the reclassification table is a  $4 \times 3$  matrix and calculation of NRI may not be straightforward. Another drawback is related to patients who experienced the event, the shift from low to intermediate or high risk groups are equally weighted. Furthermore, for those who did not experience the event, the shift from intermediate and high risk groups to low risk one is equally weighted. Furthermore, the cost of false positive equals the cost of false negative classifications. We believe that there is room to extend NRI index so as to reach a better stance that incorporate these issues.

One of the strengths of our analysis was using different statistics to explore the importance and added value of a new risk factor. We used statistics that measure the importance from likelihood goodness of fit, separation and prospect. Our study had also some limitations. First, we used a follow up data set, so future studies for case control data and for linear outcomes are necessary. Second, the majority of our variables had a binary nature, therefore our result should be checked in other datasets having categorical, binary, and continuous variables.

If we aim to fit a model, that best separate low and high risk patients, use of stepwise variable selection methods might not be optimum. This is because such approaches retain the variables that reach significance level of 0.05 rather than those that best separate the patients.

In this study, we used 7 statistics to address how definition of 'importance' affects selection of variables. We observed that a variable which its deletion does not change

the likelihood of a model, might be useful in risk stratification of the patients. Therefore, that importance should be defined based on clinical objectives of the study.

The main message of our finding is that selection of a variable in a multifactorial model should be based on the objective of the study, i.e. the comparison of a model with the highest C index or with highest separation is not necessarily the same as the model with the highest likelihood.

## Acknowledgments

The authors wish to thank Imam Reza Hospital staff in Mashhad, for their assistance.

## Footnote

**Authors' Contribution:** Study concept and design: Mohammad Reza Baneshi; acquisition of data: Ehsan Mosa Farkhani; analysis and interpretation of data: Mohammad Reza Baneshi and Saiedeh Haji-Maghsoudi; drafting of the manuscript: Mohammad Reza Baneshi, Saiedeh Haji-Maghsoudi and Ehsan Mosa Farkhani; critical revision of the manuscript for important intellectual content: Mohammad Reza Baneshi; statistical analysis: Mohammad Reza Baneshi and Saiedeh Haji-Maghsoudi; administrative, technical, and material support: Mohammad Reza Baneshi, Saiedeh Haji-Maghsoudi and Ehsan Mosa Farkhani; and study supervision: Mohammad Reza Baneshi, Saiedeh Haji-Maghsoudi and Ehsan Mosa Farkhani.

## References

- Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;**27**(2):157-72. doi: 10.1002/sim.2929. [PubMed: 17569110]
- Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;**98**(9):683-90. doi: 10.1136/heartjnl-2011-301246. [PubMed: 22397945]
- Pencina MJ, D'Agostino RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;**30**(1):11-21. doi: 10.1002/sim.4085. [PubMed: 21204120]
- Hoyt WT, Leierer S, Millington MJ. Analysis and Interpretation of Findings Using Multiple Regression Techniques. *Rehabil Couns Bull*. 2006;**49**(4):223-33. doi: 10.1177/0034352060490040401.
- Mac Nally R. Multiple regression and inference in ecology and conservation biology: further comments on identifying important predictor variables. *Biodivers Conserv*. 2002;**11**(8):1397-401. doi: 10.1023/a:1016250716679.
- Nathans LL, Oswald FL, Nimon K. Interpreting multiple linear regression: A guidebook of variable importance. 2012;**17**(9)
- Sauerbrei W, Royston P, Look M. A new proposal for multivariable modelling of time-varying effects in survival data based on fractional polynomial time-transformation. *Biom J*. 2007;**49**(3):453-73. doi: 10.1002/bimj.200610328. [PubMed: 17623349]
- Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med*. 2007;**26**(30):5512-28. doi: 10.1002/sim.3148. [PubMed: 18058845]
- Levy AR, Tamblyn RM, Fitchett D, McLeod PJ, Hanley JA. Coding accuracy of hospital discharge data for elderly survivors of myocardial infarction. *Can J Cardiol*. 1999;**15**(11):1277-82. [PubMed: 10579743]

10. Farkhani EM, Baneshi MR, Zolala F. Survival Rate And Its Related Factors In Patients With Acute Myocardial Infarction. *Med J Mashhad Univ Med Sci.* 2014;**57**(4):636-46.
11. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med.* 1999;**130**(6):515-24. [PubMed: 10075620]
12. Harrell FE, Lee KL, Mark DB. Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. *Stat Med.* 1996;**15**(4):361-87. doi: 10.1002/(sici)1097-0258(19960229)15:4<361::aid-sim168>3.0.co;2-4. [PubMed: 8668867]
13. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med.* 2004;**23**(13):2109-23. doi: 10.1002/sim.1802. [PubMed: 15211606]
14. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation.* 2007;**115**(7):928-35. doi: 10.1161/CIRCULATIONAHA.106.672402. [PubMed: 17309939]
15. Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis.* New York: Springer; 2001.
16. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem.* 2008;**54**(1):17-23. doi: 10.1373/clinchem.2007.096529. [PubMed: 18024533]
17. Greenland P, O'Malley PG. When is a new prediction marker useful? A consideration of lipoprotein-associated phospholipase A2 and C-reactive protein for stroke risk. *Arch Intern Med.* 2005;**165**(21):2454-6. doi: 10.1001/archinte.165.21.2454. [PubMed: 16314539]
18. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol.* 2004;**159**(9):882-90. [PubMed: 15105181]
19. Ware JH. The limitations of risk factors as prognostic tools. *N Engl J Med.* 2006;**355**(25):2615-7. doi: 10.1056/NEJMp068249. [PubMed: 17182986]
20. Wang TJ, Gona P, Larson MG, Tofler GH, Levy D, Newton-Cheh C, et al. Multiple biomarkers for the prediction of first major cardiovascular events and death. *N Engl J Med.* 2006;**355**(25):2631-9. doi: 10.1056/NEJMoa055373. [PubMed: 17182988]
21. Uno H, Tian L, Cai T, Kohane IS, Wei LJ. Comparing risk scoring systems beyond the ROC paradigm in survival analysis. *Harvard Univ Biostatistics Work Pap Ser.* 2009.
22. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med.* 2008;**359**(21):2208-19. doi: 10.1056/NEJMoa0804742. [PubMed: 19020323]