ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

DOI: 10.5336/biostatic.2019-71294

Evaluating the Performance of the Time Dependent Net Reclassification Improvement Under Different Settings

Farklı Düzenlerde Zamana Bağlı Net Yeniden Sınıflandırma İyileştirme Performansının Değerlendirilmesi

Eda KARAİSMAİLOĞLU[a]

[a]Kastamonu University Faculty of Medicine, Department of Biostatistics, Kastamonu, TURKEY

ABSTRACT Objective: In recent years, new measures have been proposed to evaluate the improvement in classification performance by the addition of a new risk factor to a baseline risk model that includes a set of baseline risk factors. Therefore, net reclassification improvement (NRI) and integrated discrimination improvement (IDI) methods have been utilized in medicine and these metrics have been adapted to time-to-event data in recent years. The aim of this study is to evaluate the performance of the time dependent NRI (NRI(t)) under different scenarios. **Material and Methods:** Various datasets were composed according to the purpose of each different scenario which were censoring rates (20%, 40%, and 60%), sample sizes (30, 50, 250, 500, and 1000) and number of risk categories (2, 3, and 4). Also, follow-up time was generated from Weibull distribution. All analyses and data generation process were performed using R version 3.4.3. **Results:** When number of risk categories was specified as three or four, the performance of improved model was better than two-category version. As censoring rate increased, the performance of improved model was decreased. Also, as expected, the performance of the model improved as sample size increased. In general, NRI(t) values were stable for two-category version independently of sample size and censoring rate through follow-up times. But especially for large sample sizes, the performance was higher in early time for three or four risk categories. **Conclusion:** In this study, it was found that as censoring rate decreased and number of risk categories and sample size increased, the NRI(t) improved.

Keywords: Risk prediction models; net reclassification improvement; simulation; risk factor; reclassification

ÖZET Amaç: Son yıllarda, çeşitli risk faktörlerini içeren temel risk modeline yeni bir risk faktörü eklendiğinde, sınıflandırma performansındaki iyileşmeyi değerlendirmek için yeni ölçüler önerilmiştir. Bu kapsamda, net yeniden sınıflandırma iyileştirmesi (NYSİ) ve birleştirilmiş ayrımsama iyileştirmesi (BAİ) yöntemleri tıpta kullanılmakta olup son yıllarda ilgili ölçüler sağkalım verisine de uyarlanmıştır. Bu çalışmanın amacı, zamana bağlı NRI'nın (NRI(t)) performansını farklı senaryolar altında değerlendirmektir. **Gereç ve Yöntemler:** Her bir senaryonun amacına yönelik olarak, farklı sansürleme oranlarında (% 20, % 40 ve % 60), farklı örneklem büyüklüklerinde (30, 50, 250, 500 ve 1000) ve farklı risk kategorilerinde (2, 3 ve 4) çeşitli veri setleri oluşturulmuştur. Ayrıca, takip süresi Weibull dağılımından üretilmiştir. Tüm analizler ve veri üretme süreci R programı 3.4.3 versiyonu kullanılarak yapılmıştır. **Bulgular:** Risk kategori sayısı üç veya dört olarak belirlendiğinde, iyileşmiş modelin performansı iki kategorili versiyondan daha iyidir. Sansür oranı arttıkça, iyileşmiş modelin performansı düşmüştür. Ayrıca, beklendiği gibi, örneklem büyüklüğü arttıkça iyileşmiş modelin performansı da artmıştır. Genel olarak, iki kategorili NRI(t) değerleri, örneklem büyüklüğünden ve sansürleme oranından bağımsız olarak izlem süresi boyunca değişmeyen bir performans göstermiştir. Fakat özellikle büyük örneklem büyüklüklerinde, risk kategori sayısı üç veya dört olarak alındığında, performans erken dönemde daha yüksektir. **Sonuç:** Bu çalışmada, sansürleme oranı azaldıkça ve risk kategori sayısı ve örneklem büyüklüğü arttıkça, NRI(t)'nin arttığı tespit edilmiştir.

Anahtar Kelimeler: Risk tahmin modelleri; net yeniden sınıflandırma iyileştirmesi; benzetim; risk faktörü; tekrar sınıflama

Modelling of diseases has recently been used in many areas, especially in oncology and cardiology. The main purpose of the modelling is to calculate a patient's risk (recurrence, disease, death etc.) from multivariable

Correspondence: Eda KARAİSMAİLOĞLU
Kastamonu University Faculty of Medicine, Department of Biostatistics, Kastamonu, TURKEY/TÜRKİYE
E-mail: ekaraismailoglu@kastamonu.edu.tr

Peer review under responsibility of Turkiye Klinikleri Journal of Biostatistics.

Received: 03 Sep 2019 ***Received in revised form:*** 25 Nov 2019 ***Accepted:*** 27 Nov 2019 ***Available online:*** 10 Dec 2019

models which are composed of independent variables such as biomarkers, demographic features of patients, concomitant treatment. Several statistical methods (e.g. logistic regression, Cox regression) are generally used to construct the models and a risk score for each patient is obtained through the model. It is essential to specify the disease risk with not only minimum variables (risk factors) but also optimal accuracy. In this case, the problem of interest is to determine whether a risk factor improves the predictive power of the baseline (old or existing) risk model or adds a discriminative value to the baseline model.

Area under the receiver operating characteristic curve (AUC) is widely used for evaluating incremental value of the improved (new) model which added a new risk factor to the baseline model.[1] The contribution of a new risk factor can be assessed by taking the differences of areas under the curves from the obtained baseline model (without the new risk factor) and improved model (with the new risk factor). Recently, it has been indicated that AUC is a conservative measure to capture incremental improvement. That is, improvement can be seen for risk factors that has a high relative risk or odds ratio for the disease.[2-5] In this case, new measures are needed which are more sensitive to change than AUC. For this reason, firstly, Pencina et al. (2008) suggested net reclassification improvement (NRI) and integrated discrimination improvement (IDI) methods.[6] In recent years, especially NRI has been frequently used due to its easy calculation and interpretation. Calculation of NRI is based on the classification of predicted probabilities which are obtained from the baseline model and the improved model into clinically meaningful ordinal categories of risk, which are cross-tabulated in a reclassification table. The reclassification of individuals who are in case and control groups should be considered separately. Any upward movement in categories for a case implies better classification whereas any downward movement implies worse reclassification or vice versa for a control. When the new risk factor is added, the numbers in the upper triangle are expected to be increased with respect to lower triangle for the cases and vice versa in the control group. This method shows that the degree to which a model of interest more accurately classifies people into higher or lower risk categories relative to a baseline model.[6] Meaningful cut-off values related to disease of interest is needed for the calculation of NRI. The main reason why this method has been mostly used especially in cardiovascular diseases is that guidelines recommend cut-offs for classifying individuals.[7-10]

Classical NRI does not consider follow up time but most disease status can change through time and disease result may be censored. In this case, NRI must be time dependent to eliminate bias. Pencina et al. have extended NRI method for use in survival data which is time dependent.[11] The performance of time dependent NRI (NRI(t)) may be affected by some settings. In this study, performance of NRI(t) was utilized according to different censoring rates, number of groups and cut-off values to be used for classification of individuals, and sample sizes. To address this aim, an extensive set of Monte Carlo simulations reflecting scenarios were implemented.

MATERIAL AND METHODS

Standard NRI can be calculated as below:

$$NRI = \underbrace{[P(up \mid D=1) - P(down \mid D=1)]}_{\text{NRI for the case group}} + \underbrace{[P(down \mid D=0) - P(up \mid D=0)]}_{\text{NRI for the control group}} \quad (1)$$

In the formula, D=1 indicates the case, D=0 indicates the control.

P (up | D=1), denotes the ratio of patients who pass to an upper category to all patients when a new risk factor is added to the baseline model in the cases.

P (down | D=1), denotes the ratio of patients who pass to a lower category to all patients when a new risk factor is added to the baseline model in the cases.

P (down | D=0), denotes the ratio of healthy individuals who pass to a lower category to all healthy individuals when a new risk factor is added to the baseline model in the controls.

P (up | D=0), denotes the ratio of healthy individuals who pass to an upper category to all healthy individuals when a new risk factor is added to the baseline model in the controls.

The left side of the equation can be used to calculate the NRI value for cases or patients, and the right side can be used to calculate the NRI value for healthy individuals. Thus, it can be determined whether the improved model with a newly-added risk factor categorize patients or healthy individuals better or not. Total NRI is calculated by summing up NRI values for the cases and the controls. The performance of the improved model in separating the patients and the healthy individuals is examined with total NRI. The total NRI is not a ratio and it can vary between -2 and 2.[12] Negative NRI means that the contribution of the new risk factor to the baseline model is worse and positive NRI means the contribution of the new risk factor is better.

Let M denote an estimated composite biomarker/risk score as a weighted combination of biomarkers, demographic features, and clinical variables. Higher values of M indicate a higher risk of failure (death, recurrence etc.). The estimated regression coefficients by using a Cox proportional hazards regression model can be used to obtain the composite biomarker.

$$M = \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_p X_p + \hat{\gamma}_1 Y_1 + \hat{\gamma}_2 Y_2 \cdots + \hat{\gamma}_q Y_q \quad (2)$$

In the formula, β is the regression parameters with 1xp dimensional that correspond to the biomarkers and γ is the regression parameters with qx1 dimensional that correspond to demographic features or clinical variables. The predicted survival probabilities at each observed follow-up time is obtained from the survival function of Cox proportional hazards regression model:

$$\hat{S}(t/X) = \hat{S}_0(t)^{\exp(M)} \quad (3)$$

where M is the risk score, t is follow-up time, and $S_0(t)$ is baseline hazard function. Two Cox models are fitted, one without and one with the new risk factor to obtain the predicted survival probabilities. After that, the predicted probabilities are categorized according to clinically meaningful cut-off values for both models.

Let D(t) denote status of an event's occurrence and T denote failure time. If $T_j \leq t$, then D(t)=1 which indicates the event and if $T_j > t$, then D(t)=0 which indicates censored for j. individual. For survival outcomes, Kaplan Meier estimator (KM) can be used for estimating the number of cases and controls in the cross table obtained by categorizing the predicted probabilities at time t. Thus for censored survival outcomes, NRI at time t can be calculated for the estimated number of cases and controls *moving up* or *moving down* risk categories.[11] To make it more clear, NRI for 3x3 table was summarized in Table 1. Rows indicate the baseline model's risk categories and columns indicate the improved model's (with added new risk factor) risk categories in the table.

NRI(t) is obtained by the following equation.

TABLE 1: Reclassification table for NRI(t) calculation with given risk categories.

	Group	Improved Model		
		Low Risk	**Moderate Risk**	**High Risk**
Baseline Model	**Low Risk**	-	**Up** Improves when D(t)=1 Worsens when D(t)=0	**Up** Improves when D(t)=1 Worsens when D(t)=0
	Moderate Risk	**Down** Worsens when D(t)=1 Improves when D(t)=0	-	**Up** Improves when D(t)=1 Worsens when D(t)=0
	High Risk	**Down** Worsens when D(t)=1 Improves when D(t)=0	**Down** Worsens when D(t)=1 Improves when D(t)=0	-

$$NRI(t)=P\ up\ D\ t=1\ -P\ (down\mid D\ (t)=1)]+[P\ (down\ \ D\ t=0-P\ up\ D\ t=0)] \quad (4)$$

The equation is based on Bayes theorem and survival function. The first part of the formulation expresses the probability of moving up for the cases at time t and the second part expresses the probability of moving down for the controls at time t. Total NRI(t) is equal to summation of NRI(t) case and control.[11,13] Expanding the components of the total NRI(t):

$$P(up|D(t) = 1) = \frac{(1-S(t|up)P(up))}{(1-S(t))} \quad (5)$$

$$P(down|D(t)=1)= \frac{(1-S(t|down)P(down))}{(1-S(t))} \quad (6)$$

$$P(up|D(t) = 0) = \frac{S(t|up)P(up)}{S(t)} \quad (7)$$

$$P(down|D(t) = 0) = \frac{(St|down)P(down)}{S(t)} \quad (8)$$

By Bayes' theorem, the estimator of NRI (t) can be obtained as follows:

$$NRI(t)=[P\ (up\mid D(t)=1)-P\ (down\mid D(t)=1)]+[P\ (down\mid D(t)=0)-P\ (up\mid D(t)=0)] \quad (9)$$

Expansion of equation (9) is as follows:

$$\hat{P}(up|D(t) = 1) = \frac{(\hat{P}(D(t) =1|up).\hat{P}(up)}{\hat{P}(D(t)=1)} \quad (10);\ \hat{P}(down|D(t) = 1) = \frac{(\hat{P}(D(t) =1\ |down).\hat{P}(down)}{\hat{P}(D(t)=1)} \quad (11)$$

$$\hat{P}(up|D(t) = 0) = \frac{(\hat{P}(D(t) =0\ |up).\hat{P}(up)}{\hat{P}(D(t)=0)} \quad (12);\ \hat{P}(down|D(t) = 0) = \frac{(\hat{P}(D(t) =0|down).\hat{P}(down)}{\hat{P}(D(t)=0)} \quad (13)$$

where P(down) and P(up) are calculated as follows

$$\hat{P}(down) = \frac{\text{Number of individuals moving down}}{\text{Total individuals}} \quad (14);$$

$$\hat{P}(up) = \frac{\text{Number of individuals moving up}}{\text{Total Individuals}} \quad (15)$$

Kaplan Meier method can be used to estimate P(D(t)=1|up), P(D(t)=1|down), P(D(t)=0|up), P(D(t)=0|down). Function can be estimated at t time as follows.[14]

Let event times are $t_1 \leq t_2 \leq \ldots \leq t_n$

$$\hat{S(t)} = \prod tk \leq t\ \frac{n_k - d_k}{n_k} \quad (16)$$

n_k is number of individuals who are at risk at t_k time and d_k is number of event at t_k time.

Monte Carlo Simulations

In this section, a simulation study was carried out to investigate the influence of different scenarios on NRI(t) measure.

x_1 and x_2 risk factors were created from a standard bivariate normal distribution with 0.3 correlation and different sample sizes between 30 and 1000.

$$X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix},\begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}\right) \quad (17)$$

To generate survival times for each individual, Weibull distribution which guarantees a proportional hazards relationship was used, characterized by shape parameter λ and scale parameter γ. The formula is as follows,

$$T = \left(-\frac{\log(U)}{\gamma \exp(\beta x)}\right)^{1/\lambda} \quad (18)$$

where U ~ Uniform (0,1), x vector of the risk factors, β vector of the regression coefficients. As generating survival times, shape parameter was set as 1 and scale parameter was set as exp $(-\beta_1 x_1 - \beta_2 x_2)$. Similarly, censoring times were generated with shape parameter that was 1 and scale parameter that differed according to censoring rates (for 60% censoring rate, γ=0.09; for 40% censoring rate, γ=0.05; for 20% censoring rate, γ=0.01). Observed follow-up time was defined as $t=\min(t_{event}, t_{censored})$. Follow-up time period was specified as [0, 2]. The relationship between survival time T and X fit in a proportional hazards model with parameters $\beta_1 = -1.5$ and $\beta_2 = 2.8$.

Risk categories for NRI(t) should be determined according to clinically meaningful cut-off values. For this reason, cut-off values were determined according to the literature that are generally recommended for primary prevention of cardiovascular disease.[7-10]

In the first step of the simulation, data were generated consisting of biomarkers or risk factors, survival times and status of disease according to different settings. In the second step, Cox proportional hazards model was composed to obtain risk scores of individuals for both the baseline and the improved models. Risk scores were obtained from linear predictors as follows.

Risk scores for the baseline model $RS_{baseline}=\hat{\beta}_1 x_1$

Risk scores for the improved model $RS_{improved}=\hat{\beta}_1 x_1+\hat{\beta}_2 x_2$

The baseline model was composed with x_1 and the improved model was composed by adding x_2 to the baseline model. In the third step, predicted survival probabilities were obtained by using risk scores at each observed time. In the fourth step, predicted probabilities were categorized according to specified cut-offs. Afterwards, Kaplan-Meier method was used to estimate cases and controls. Lastly, NRI(t) value was obtained by using equations (10 – 13).

The performance of NRI(t) was evaluated according to

(i) different sample sizes which were 30, 50, 250, 500, and 1000,

(ii) different right censoring rates which were 20%, 40%, and 60%,

(iii) different number of categories which were two groups (cut-off: 50%), three groups (cut-offs: 20% and 50%), and four groups (cut-offs: 5%, 10%, and 20%) on the NRI(t).

Thus, 45 (5×3×3) different scenarios were examined. All scenarios were repeated 1000 times and results were recorded as a mean of NRI(t) for each observed time. All simulations were implemented with R software version 3.4.3- (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Results were reported separately according to the changes in censoring rates, sample sizes, and number of categories (Figures 1, 2, and 3 respectively). In all settings, the newly added risk factor x2 improved the baseline model more or less. Since NRI(t)'s values were greater than zero, y axes of graphs were determined as [0,2].

In all sample sizes, the performance of improved (new) model decreased as censoring rate increased. In other words, while censoring rate increased (or event rate decreased), there was a decrease or no improvement in the performance of the improved model after adding the new risk factor. On the other hand, as number of categories (number of groups) increased, the performance of improved model increased. The highest performance was observed at 20% censoring rate when sample size was greater than 500 and number of categories was three or four. In the same settings, there was a higher performance during early times although the performance started to decrease later on. On the other hand, there was not any trend over time in small sample sizes (Figure 1).

In general, as sample sizes increased, the performance of improved model increased independently of censoring rate and number of categories but performance change was more stable when number of categories was specified as two regardless of the censoring rate or sample size. When number of categories was more than two and sample size was more than 500, the performance of improved model showed dramatic enhancement in 20% censoring rate. But the performance showed decreasing trend through follow-up time (Figure 2).

The performance of improved model was close to each other independently of censoring rate and sample size when number of categories was three or four. Similarly, three and four categories were better than two categories in all settings. NRI(t) values were more unstable in these categories during follow-up time. The difference between the two and other categories (3 and 4) was remarkable especially in large sample sizes (greater than 250) (Figure 3).

The performance of the NRI(t) was presented in terms of bias for t=0.5, 1, 1.5, and 2 time points which were divided into equal parts. It can be observed that the biases were close to zero for these time points (Table 2-6). Moreover, the overall performance of the NRI(t) can be seen for each observed time points in the graphs. The real NRI(t) was indicated by dotted line and the mean of bootstrap NRI(t) was indicated by solid line in the graphs. The biases were close to zero as such in specified time points (Figure 1-3).

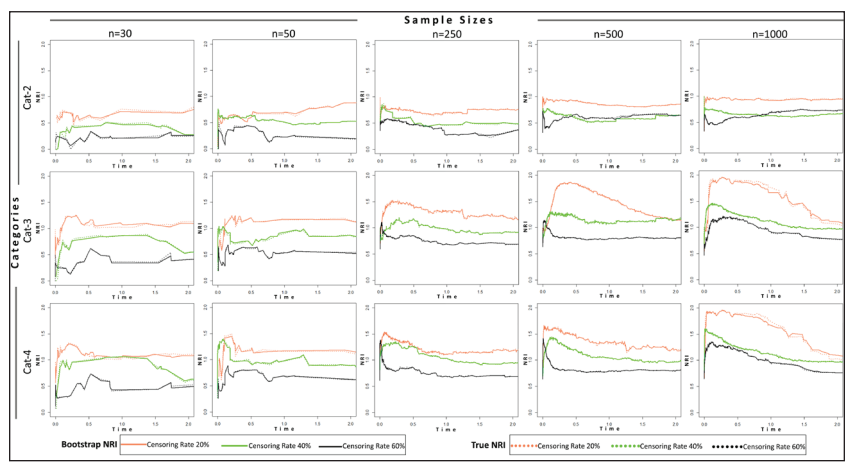

FIGURE 1: True and bootstrap NRI(t) values according to censoring rates (20%, 40%, and 60%) for different categories (2, 3, and 4) and different sample sizes (30, 50, 250, 500, and 1000).
Cat: category.

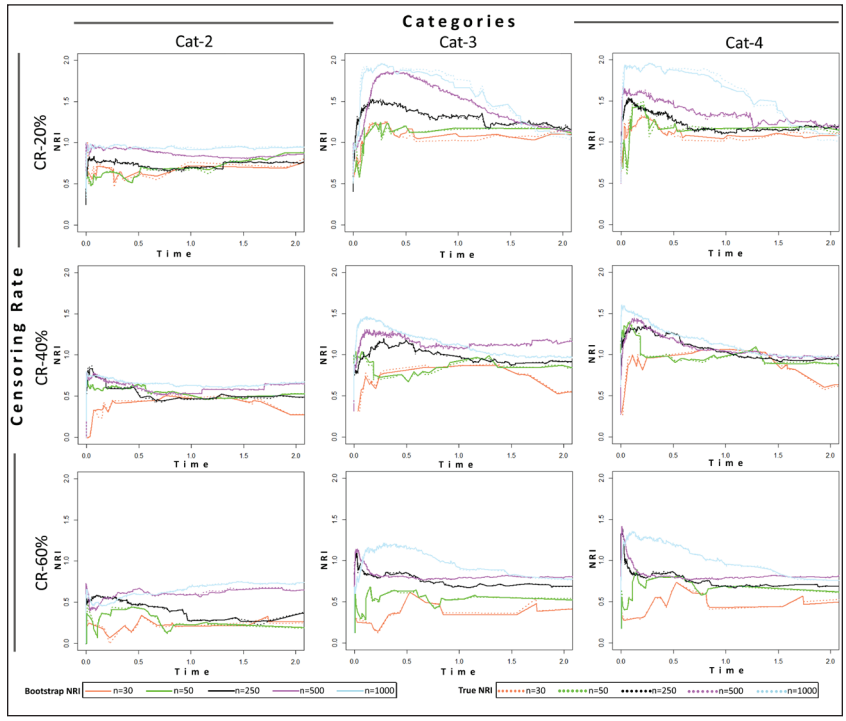

FIGURE 2: True and bootstrap NRI(t) values according to sample sizes (30, 50, 250, 500, and 1000) for different categories (2, 3, and 4) and different censoring rates (20%, 40%, and 60%).
CR: censoring rate; Cat: category.

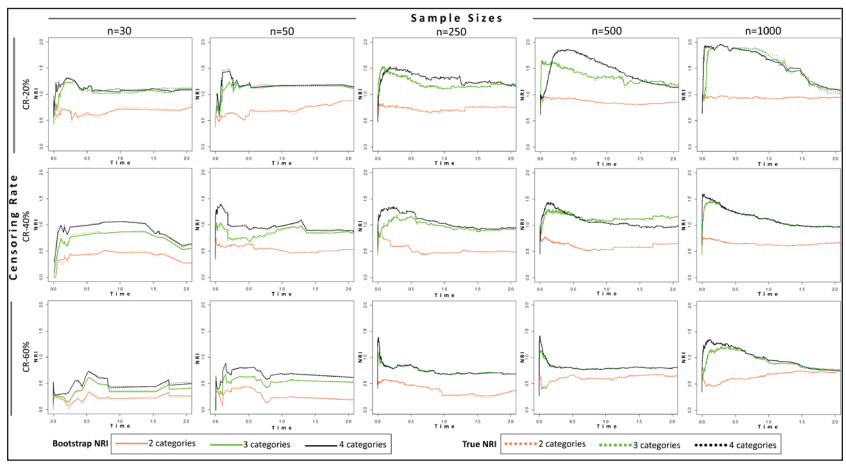

FIGURE 3: True and bootstrap NRI(t) values according to categories (2, 3, and 4) for different censoring rates (20%, 40%, and 60%) and different sample sizes (30, 50, 250, 500, and 1000).
CR: censoring rate.

TABLE 2: True NRI(t) and mean of bootstrap NRI(t) from 1000 simulations (n=30) at time points 0.5, 1, 1.5 and 2.

CR	Time	Cat - 2			Cat - 3			Cat - 4		
		True NRI(t)	Bootstrap NRI(t)	Bias	True NRI(t)	Bootstrap NRI(t)	Bias	True NRI(t)	Bootstrap NRI(t)	Bias
20%	0.5	0.620	0.648	-0.028	0.516	0.510	0.006	1.064	1.099	-0.035
	1	0.764	0.720	0.044	0.363	0.348	0.015	1.054	1.084	-0.030
	1.5	0.718	0.714	0.004	0.361	0.342	0.019	1.095	1.107	-0.012
	2	0.724	0.740	-0.016	0.425	0.425	0	1.099	1.078	0.021
40%	0.5	0.441	0.416	0.025	0.809	0.806	0.003	1.030	1.018	0.012
	1	0.482	0.469	0.013	0.868	0.869	-0.001	1.061	1.065	-0.004
	1.5	0.431	0.430	0.001	0.773	0.778	-0.005	0.860	0.826	0.034
	2	0.276	0.278	-0.002	0.556	0.546	0.010	0.611	0.633	-0.022
60%	0.5	0.216	0.238	-0.022	1.048	1.117	-0.069	0.633	0.628	0.005
	1	0.200	0.208	-0.008	1.020	1.069	-0.049	0.447	0.436	0.011
	1.5	0.253	0.220	0.033	1.058	1.071	-0.013	0.444	0.442	0.002
	2	0.250	0.264	-0.014	1.131	1.102	0.029	0.533	0.504	0.029

CR: censoring rate; Cat: category.

TABLE 3: True NRI(t) and mean of bootstrap NRI(t) from 1000 simulations (n=30) at time points 0.5, 1, 1.5 and 2.

CR	Time	Cat - 2			Cat - 3			Cat - 4		
		True NRI(t)	Bootstrap NRI(t)	Bias	True NRI(t)	Bootstrap NRI(t)	Bias	True NRI(t)	Bootstrap NRI(t)	Bias
20%	0.5	0.590	0.604	-0.014	1.209	1.203	0.006	1.207	1.175	0.032
	1	0.700	0.702	-0.002	1.156	1.176	-0.020	1.156	1.168	-0.012
	1.5	0.775	0.775	0	1.161	1.179	-0.018	1.156	1.173	-0.017
	2	0.870	0.873	-0.003	1.048	1.083	-0.035	1.054	1.098	-0.044
40%	0.5	0.638	0.641	-0.003	0.678	0.670	0.008	0.922	0.906	0.016
	1	0.504	0.507	-0.003	0.907	0.916	-0.009	0.966	0.980	-0.014
	1.5	0.477	0.474	0.003	0.858	0.842	0.016	0.900	0.897	0.003
	2	0.530	0.530	0	0.870	0.872	-0.002	0.894	0.884	0.010
60%	0.5	0.417	0.418	-0.001	0.606	0.629	-0.023	0.801	0.809	-0.008
	1	0.224	0.225	-0.001	0.526	0.529	-0.003	0.680	0.671	0.009
	1.5	0.227	0.236	-0.009	0.553	0.565	-0.012	0.677	0.687	-0.010
	2	0.217	0.228	-0.011	0.522	0.528	-0.006	0.667	0.670	-0.003

CR: censoring rate; Cat: category.

TABLE 4: True NRI(t) and mean of bootstrap NRI(t) from 1000 simulations (n=250) at time points 0.5, 1, 1.5 and 2.

CR	Time	Cat - 2			Cat - 3			Cat - 4		
		True NRI(t)	Bootstrap NRI(t)	Bias	True NRI(t)	Bootstrap NRI(t)	Bias	True NRI(t)	Bootstrap NRI(t)	Bias
20%	0.5	0.700	0.709	-0.009	1.370	1.383	-0.013	1.259	1.251	0.008
	1	0.685	0.700	-0.015	1.288	1.305	-0.017	1.120	1.118	0.002
	1.5	0.759	0.758	0.001	1.220	1.229	-0.009	1.134	1.132	0.002
	2	0.753	0.756	-0.003	1.167	1.160	0.007	1.212	1.198	0.014
40%	0.5	0.490	0.481	0.009	1.169	1.177	-0.008	1.270	1.261	0.009
	1	0.461	0.473	-0.012	0.982	0.974	0.008	1.021	1.030	-0.009
	1.5	0.488	0.498	-0.010	0.872	0.875	-0.003	0.911	0.912	-0.001
	2	0.484	0.485	-0.001	0.918	0.915	0.003	0.955	0.955	0
60%	0.5	0.406	0.466	-0.006	0.838	0.847	-0.009	0.853	0.849	0.004
	1	0.270	0.271	-0.001	0.685	0.689	-0.004	0.676	0.686	-0.010
	1.5	0.269	0.272	-0.003	0.688	0.690	-0.002	0.679	0.685	-0.006
	2	0.376	0.372	0.004	0.687	0.691	-0.004	0.687	0.688	-0.001

CR: censoring rate; Cat: category.

TABLE 5: True NRI(t) and mean of bootstrap NRI(t) from 1000 simulations (n=500) at time points 0.5, 1, 1.5 and 2.

CR	Time	Cat - 2			Cat - 3			Cat - 4		
		True NRI(t)	Bootstrap NRI(t)	Bias	True NRI(t)	Bootstrap NRI(t)	Bias	True NRI(t)	Bootstrap NRI(t)	Bias
20%	0.5	0.922	0.921	0.001	1.841	1.838	0.003	1.418	1.424	-0.006
	1	0.833	0.835	-0.002	1.545	1.540	0.005	1.331	1.339	-0.008
	1.5	0.812	0.814	-0.002	1.269	1.272	-0.003	1.244	1.241	0.003
	2	0.854	0.857	-0.003	1.141	1.140	0.001	1.198	1.205	-0.007
40%	0.5	0.594	0.599	-0.005	1.195	1.202	-0.007	1.189	1.204	-0.015
	1	0.526	0.528	-0.002	1.078	1.078	0	1.046	1.042	0.004
	1.5	0.578	0.582	-0.004	1.121	1.125	-0.004	0.984	0.981	0.003
	2	0.650	0.652	-0.002	1.164	1.161	0.003	0.976	0.963	0.013
60%	0.5	0.662	0.669	-0.007	0.796	0.793	0.003	0.790	0.789	0.001
	1	0.584	0.588	-0.004	0.795	0.790	0.005	0.794	0.793	0.001
	1.5	0.658	0.652	0.006	0.812	0.811	0.001	0.817	0.813	0.004
	2	0.642	0.635	0.007	0.805	0.799	0.006	0.802	0.798	0.004

CR: censoring rate; Cat: category.

TABLE 6: True NRI(t) and mean of bootstrap NRI(t) from 1000 simulations (n=1000) at time points 0.5, 1, 1.5 and 2.

CR	Time	Cat - 2			Cat - 3			Cat - 4		
		True NRI(t)	Bootstrap NRI(t)	Bias	True NRI(t)	Bootstrap NRI(t)	Bias	True NRI(t)	Bootstrap NRI(t)	Bias
20%	0.5	0.937	0.935	0.002	1.884	1.866	0.018	1.884	1.871	0.013
	1	0.922	0.919	0.003	1.724	1.699	0.025	1.701	1.683	0.018
	1.5	0.953	0.950	0.003	1.398	1.411	-0.013	1.358	1.377	-0.019
	2	0.952	0.948	0.004	1.059	1.083	-0.024	1.083	1.095	-0.012
40%	0.5	0.652	0.655	-0.003	1.234	1.247	-0.013	1.250	1.262	-0.012
	1	0.627	0.629	-0.002	1.137	1.138	-0.001	1.111	1.114	-0.003
	1.5	0.617	0.621	-0.004	0.983	0.989	-0.006	0.985	0.989	-0.004
	2	0.670	0.664	0.006	0.970	0.964	0.006	0.969	0.962	0.007
60%	0.5	0.599	0.594	0.005	1.139	1.144	-0.005	1.177	1.191	-0.014
	1	0.683	0.691	-0.008	0.905	0.920	-0.015	0.928	0.944	-0.016
	1.5	0.748	0.741	0.007	0.828	0.819	-0.009	0.809	0.831	-0.022
	2	0.729	0.732	-0.003	0.766	0.773	-0.007	0.754	0.759	-0.005

CR: censoring rate; Cat: category.

DISCUSSION

Recently, NRI measure has been very popular to evaluate the contribution of the newly-added biomarker or risk factor to a baseline risk prediction model. It is based on movement in predicted probabilities of an event or a control between clinically meaningful risk categories. It evaluates whether adding a new risk factor to the baseline model improves the number of individuals with higher predicted probabilities among events and decreases the number of individuals with higher predicted probabilities among controls.[11] The NRI is generally applied to the nested models, where the baseline model is a subset of the risk categories in the improved model. On the other hand, NRI(t) has been developed to utilize survival data but there is a limited literature on NRI(t).[11,12,15-18] Some of those proposed a new method to estimate NRI(t).[15,17] Liu et al. compared performances of standard NRI and NRI(t).[16] French et al. modelled biomarkers with quadratic and interaction forms and compared the performances of the models with NRI(t).[18] To the best of our knowledge, this simulation study is the first to compare NRI(t) under different scenarios which included a variety of sample sizes, censoring rates and number of categories. Mühlenbruch et al. analyzed the impact of number of risk categories and risk cut-offs over standard NRI.[19] They found that NRI value improved with increasing numbers of categories but this increment was not monotone. However, Pencina et al. did not recommend using more than three catego-

ries.[11] We showed that improved model exhibited better performance with three and four categories than two categories. In addition, three and four categories showed similar performance regardless of sample size and censoring rate. For this reason, we also suggest that using three categories is preferable to estimate NRI(t). NRI or NRI(t) should be used for events in which there are clinically meaningful risk categories with broad acceptance.[20] Alternatively, Pencina et al. recommended that event rate can be accepted for the two-category version of NRI in cases where there are not clinically meaningful risk categories.[21] A limitation of this study is that the choice of cut-off points may change the results of NRI(t). Therefore, we preferred the most frequently used cut-offs in the literature which are used for primary prevention of cardiovascular disease. Another study can be planned to evaluate change of the NRI(t) according to different cut-off values.

CONCLUSION

As a result, it has been found that two-category version of NRI(t) showed lower results in contrast with three and four categories in all settings as the previous studies about standard NRI. The pattern of two-category of NRI(t) for improvements in the model performance was stable through follow-up time. Therefore, it is suggested that more than two categories should be used while implementing NRI(t) as a performance measure. As sample size increases, the NRI(t) improves as expected and the more preferable results can be obtained with lower censoring rate.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and / or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

This study is entirely author's own work and no other author contribution.

REFERENCES

1. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837-45. PMID: 3203132. [Crossref] [PubMed]
2. 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. PMID: 15105181. [Crossref] [PubMed]
3. 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. PMID: 17182988. [Crossref] [PubMed]
4. Ware JH. The limitations of risk factors as prognostic tools. N Engl J Med. 2006;355(25):2615-7. PMID: 17182986. [Crossref] [PubMed]
5. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. Ann Intern Med. 2009;150(11):795-802. PMID: 19487714. [Crossref] [PubMed] [PMC]
6. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, 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. PMID: 17569110. [Crossref] [PubMed]
7. Perk J, De Backer G, Gohlke H, Graham I, Reiner Z, Verschuren M, et al; European Association for Cardiovascular Prevention & Rehabilitation (EACPR). European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). Eur Heart J. 2012;33(13):1635-701. PMID: 22555213.
8. Greenland P, Alpert JS, Beller GA, Benjamin EJ, Budoff MJ, Fayad ZA, et al; American College of Cardiology Foundation/American Heart AssociationTask Force on Practice Guidelines. 2010 ACCF/AHA guideline for assessment of cardiovascular risk in asymptomatic adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation. 2010;122(25):e584-636. PMID: 21098428. [Crossref] [PubMed]
9. Mosca L, Benjamin EJ, Berra K, Bezanson JL, Dolor RJ, Lloyd-Jones DM, et al. Effectiveness-based guidelines for the prevention of cardiovascular disease inwomen-2011 update: a guideline from the American Heart Association. Circulation.2011;123(11):1243-62. PMID: 21325087. [Crossref] [PubMed] [PMC]

10. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). JAMA. 2001;285(19):2486-97. [Crossref] [PubMed]
11. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med. 2011;30(1):11-21. PMID: 21204120. [Crossref] [PubMed] [PMC]
12. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. Epidemiology. 2014;25(1):114-21. PMID: 24240655. [Crossref] [PubMed] [PMC]
13. Viallon V, Ragusa S, Clavel-Chapelon F, Bénichou J. How to evaluate the calibration of a disease risk prediction tool? Stat Med. 2009;28(6):901-16. PMID: 19156698. [Crossref] [PubMed]
14. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958;53:457-81. [Crossref]
15. Zheng Y, Parast L, Cai T, Brown M. Evaluating incremental values from new predictors with net reclassification improvement in survival analysis. Lifetime Data Anal. 2013;19(3):350-70. PMID: 23254468. [Crossref] [PubMed] [PMC]
16. Liu M, Kapadia AS, Etzel CJ. Evaluating a new risk marker's predictive contribution in survival models. J Stat Theory Pract. 2010;4(4):845-55. PMID: 22984361. [Crossref] [PubMed] [PMC]
17. Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. Stat Med. 2011;30(1):22-38. PMID: 20827726. [Crossref] [PubMed]
18. French B, Saha-Chaudhuri P, Ky B, Cappola TP, Heagerty PJ. Development and evaluation of multi-marker risk scores for clinical prognosis. Stat Methods Med Res. 2016;25(1):255-71. PMID: 22767865. [Crossref] [PubMed] [PMC]
19. Mühlenbruch K, Heraclides A, Steyerberg EW, Joost HG, Boeing H, Schulze MB. Assessing improvement in disease prediction using net reclassification improvement: impact of risk cut-offs and number of risk categories. Eur J Epidemiol. 2013;28(1):25-33. PMID: 23179629. [Crossref] [PubMed]
20. Pickering JW, Endre ZH. New metrics for assessing diagnostic potential of candidate biomarkers. Clin J Am Soc Nephrol. 2012;7(8):1355-64. PMID: 22679181. [Crossref] [PubMed]
21. Pencina MJ, Steyerberg EW, D'Agostino RB Sr. Net reclassification index at event rate: properties and relationships. Stat Med. 2017;36(28):4455-67. PMID: 27426413. [Crossref] [PubMed]