Evaluating a New Risk Marker's Predictive Contribution in Survival Models

M. LIU,
Department of Epidemiology, University of Texas, MD Anderson Cancer Center, Houston, TX 77030

A. S. KAPADIA, and
Department of Biostatistics, University of Texas, School of Public Health, Houston, TX 77030

C. J. ETZEL
Department of Epidemiology, University of Texas, MD Anderson Cancer Center, Houston, TX 77030

M. LIU: meiliu@mdanderson.org; A. S. KAPADIA: Asha.S.Kapadia@uth.tmc.edu; C. J. ETZEL: cetzel@mdanderson.org

Abstract

Although the area under the receiver operating characteristic (ROC) curve (AUC) is the most popular measure of the performance of prediction models, it has limitations, especially when it is used to evaluate the added discrimination of a new risk marker in an existing risk model. Pencina et al. (2008) proposed two indices, the net reclassification improvement (NRI) and integrated discrimination improvement (IDI), to supplement the improvement in the AUC (IAUC). Their NRI and IDI are based on binary outcomes in case-control settings, which do not involve time-to-event outcome. However, many disease outcomes are time-dependent and the onset time can be censored. Measuring discrimination potential of a prognostic marker without considering time to event can lead to biased estimates. In this paper, we extended the NRI and IDI to time-to-event settings and derived the corresponding sample estimators and asymptotic tests. Simulation studies showed that the time-dependent NRI and IDI have better performance than Pencina's NRI and IDI for measuring the improved discriminatory power of a new risk marker in prognostic survival models.

Keywords

Improved discrimination; Prognostic survival models; Time-dependent NRI; Time-dependent IDI

1. Introduction

Risk prediction models are important tools to assess risk for cancer and other complex diseases and have wide applications in clinical medicine, clinical trials, and allocation of health services resources (Ridker et al. 2008). Calibration and discrimination are two major components in the evaluation of the predictive accuracy of a risk prediction model (Cook 2007). Calibration quantifies how closely the predicted probabilities agree numerically with the actual outcomes. Discrimination measures the ability of the model to correctly separate subjects with different outcomes. A model is said to have perfect discrimination if the model places each individual in the category to which he/she truly belongs. Good discrimination of a model does not automatically imply good calibration or vice versa (Cook 2007). Harrell (Harrell, Jr. et al. 1996) suggests discrimination is of more interest and should be the primary focus.

The area under the receiver operating characteristic (ROC) curve (AUC) is the most popular measure to evaluate model discrimination by calculating the probability that among all possible pairs of individuals with two different outcomes – positive versus negative, the predicted value for the one with positive outcome is higher than the one with negative outcome (Hanley and Mcneil 1982). Increase in AUC (IAUC), defined as the difference in the AUCs calculated using the model with and the model without the new markers of interest, is the most widely used measure to quantify the model improvement due to the addition of a new risk marker to the existing risk models. However, this increase is often very small in magnitude and insensitive in evaluating the new marker's predictive contribution, especially when a few powerful risk factors are already in the model.

Pencina et al. (Pencina et al. 2008) proposed two new measures to quantify the increase in performance when a new marker is added to an existing risk model. The first is 'net reclassification improvement' (NRI), a measure based on event-specific reclassification tables and compares the proportions moving up or down in clinical categories in cases versus controls. They define upward movement (up) as a change into higher category based on the new model and downward movement (down) as the change into lower category based on the new model. In their paper, D denotes the event (or disease) indicator ($D = 1$ for cases and $D = 0$ for controls), and the NRI is defined as

$$NRI = (P(up|D{=}1) - P(down|D{=}1)) - (P(up|D{=}0) - P(down|D{=}0)).$$

The second summary measure is the 'integrated discrimination improvement' (IDI), which examines the difference in the mean predicted risk between individuals who develop the event and those who don't. It can also be interpreted as a weighted area between the ROC curve and the diagonal line in the ROC plot. For any cut-off point u ($0 < u < 1$), the IDI is defined as

$$IDI = (IS_{new} - IS_{old}) - (IP_{new} - IP_{old}),$$

where $IS = \int_0^1 sensitivity(u)du$ and $IP = \int_0^1 (1 - specificity(u))du$. The subscript 'new' refers to the model with the new marker while 'old' refers to the original model. Both of these new measures offer incremental information over the IAUC and have proved useful for evaluating the predictive power of new risk markers (Zethelius et al. 2008; Simmons et al. 2008; Eggers et al. 2009; Shah et al. 2009).

The NRI and IDI indices proposed by Pencina et al. are based on binary outcomes in case-control settings, which do not involve time-to-event outcomes. However, many disease outcomes are time-dependent and the disease onset can be censored. Measuring discrimination potential of a marker in the context of survival analysis (time-to-event analysis) without consideration of time-to-event can lead to biased estimates (Chambless and Diao 2006). In this paper, we extended NRI and IDI to survival analysis settings where the disease onset is observed over follow-up time periods that vary in length. This extension provides a more accurate measure to evaluate the predictive ability of a new marker added to the risk of mortality or disease onset at time t.

2. Methods

2.1 Time-dependent NRI (tdNRI)

Under the survival time setting, we introduce the following notations:

For subject i, let T_i denote the failure time and C_i the censoring time, then $Y_i = \min(T_i, C_i)$ is the follow-up time. Let $D_i(t)$ represent failure status at time t, which indicates whether subject i had an event prior to time t; that is, $D_i(t) = 1$ if $T_i \leq t$ and $D_i(t) = 0$ if $T_i > t$. Z_i is a risk score that can be used to predict whether subject i would developed an event by time t.

Let X denote the vector of covariates (risk factors). In survival regression model, Z is usually calculated as the sum of products of the subject's level for each risk factor (X) and the corresponding estimated regression coefficient vector ($\hat{\beta}$), that is, $Z = X\hat{\beta}$. Z_{old} is the vector of risk scores calculated from the original model and Z_{new} as the vector of risk scores calculated from the new model. Here we assume that the new model includes all of the variables from the original model plus one new risk factor of interest. Let 'Up' and 'down' represent the events that the new model moves the predicted risk score up and down, respectively. For example, for subject i, if $(Z_{new})_i > (Z_{old})_i$, we say subject i is reclassified up by the new model.

Now we define the time-dependent NRI (tdNRI) as a function of time:

$$tdNRI(t) = (P(up|D(t)=1) - P(down|D(t)=1)) - (P(up|D(t)=0) - P(down|D(t)=0)). \quad (2.1)$$

The tdNRI for the subjects who have events by time t and for those who don't are defined respectively as follows:

$$tdNRI\_event(t) = P(up|D(t)=1) - P(down|D(t)=1), \quad (2.2)$$

$$tdNRI\_nonevent(t) = P(down|D(t)=0) - P(up|D(t)=0). \quad (2.3)$$

Using Bayes' theorem, we can rewrite $P(up|D(t) = 1)$, $P(down|D(t) = 1)$, $P(up|D(t) = 0)$, and $P(down|D(t) = 0)$ as

$$P(up|D(t)=1) = \frac{(1 - S(t|up))P(up)}{1 - S(t)}, \quad (2.4)$$

$$P(down|D(t)=1) = \frac{(1 - S(t|down))P(down)}{1 - S(t)}, \quad (2.5)$$

$$P(up|D(t)=0) = \frac{S(t|up)P(up)}{S(t)}, \quad (2.6)$$

and

$$P(down|D(t)=0) = \frac{S(t|down)P(down)}{S(t)}, \quad (2.7)$$

where $S(t)$ is the survival function $S(t) = P(T > t)$. $S(t|up)$ and $S(t|down)$ are the conditional survival functions for the subset of subjects satisfying $Z_{new} > Z_{old}$ and $Z_{new} < Z_{old}$, respectively; while $P(up)$ or $P(down)$ are the probabilities that $Z_{new} > Z_{old}$ or $Z_{new} < Z_{old}$.

The most popular nonparametric method to estimate the survival function $S(t)$ is Kaplan-Meier or product limit method (1958). Assume that the events occur at D distinct times $t_1 < t_2 < \ldots < t_D$, and at time t_i there are d_i events. Define N_i as the number of individuals who are at risk at time t_i. For any time t ($t_1 \leq t$), the Kaplan-Meier (KM) estimator $\hat{S}_{KM}(t)$ is defined as

$$\widehat{S}_{KM}(t) = \prod_{t_i \leq t} \left\{ 1 - \frac{d_i}{N_i} \right\}. \tag{2.8}$$

So a simple estimator for $P(up|D(t) = 1)$ is given by

$$\widehat{P}(up|D(t)=1) = \frac{(1 - \widehat{S}_{KM}(t|up))\widehat{P}(up)}{1 - \widehat{S}_{KM}(t)}. \tag{2.9}$$

Likewise, the following expressions are the estimators for $P(down|D(t) = 1)$, $P(up|D(t) = 0)$ and $P(down|D(t) = 0)$:

$$\widehat{P}(down|D(t)=1) = \frac{(1 - \widehat{S}_{KM}(t|down))\widehat{P}(down)}{1 - \widehat{S}_{KM}(t)}, \tag{2.10}$$

$$\widehat{P}(up|D(t)=0) = \frac{\widehat{S}_{KM}(t|up)\widehat{P}(up)}{\widehat{S}_{KM}(t)}, \tag{2.11}$$

$$\widehat{P}(down|D(t)=0) = \frac{\widehat{S}_{KM}(t|down)\widehat{P}(down)}{\widehat{S}_{KM}(t)}, \tag{2.12}$$

where

$$\widehat{P}(up) = \frac{\#subjects\ moving\ up}{total\ subjects} = \frac{\#(i|(Z_{new})_i > (Z_{old})_i)}{total\ subjects} \tag{2.13}$$

and

$$\widehat{P}(down) = \frac{\#subjects\ moving\ down}{total\ subjects} = \frac{\#(i|(Z_{new})_i < (Z_{old})_i)}{total\ subjects}. \tag{2.14}$$

Thus, $tdNRI(t)$ is estimated as

$$td\widehat{NRI}(t) = (\widehat{P}(up|D(t)=1) - \widehat{P}(down|D(t)=1)) - (\widehat{P}(up|D(t)=0) - \widehat{P}(down|D(t)=0)). \tag{2.15}$$

Similarly, in formulas (2.2) and (2.3), the tdNRI for event subjects and nonevent subjects by time t may be estimated separately as

$$tdN\widehat{RI}\_event(t) = \widehat{P}(up|D(t)=1) - \widehat{P}(down|D(t)=1),$$ (2.16)

$$tdN\widehat{RI}\_noevent(t) = \widehat{P}(down|D(t)=0) - \widehat{P}(up|D(t)=0).$$ (2.17)

Assuming that the event individuals and nonevent individuals at time t are independent, we may use a simple asymptotic test testing the null hypothesis of $tdNRI(t) = 0$, with the following test statistic:

$$z(t) = \frac{tdN\widehat{RI}(t)}{\sqrt{\frac{tdN\widehat{RI}\_event(t)(1-tdN\widehat{RI}\_event(t))}{100*(1-\widehat{S}_{KM}(t))} + \frac{tdN\widehat{RI}\_noevent(t)(1-tdN\widehat{RI}\_noevent(t))}{100*(\widehat{S}_{KM}(t))}}}.$$ (2.18)

The improvement in event and non-event classification at time t can be tested by the following formulas:

$$z_{event}(t) = \frac{\widehat{P}(up|D(t)=1) - \widehat{P}(down|D(t)=1)}{\sqrt{\frac{\widehat{P}(up|D(t)=1)(1-\widehat{P}(up|D(t)=1))+\widehat{P}(down|D(t)=1)(1-\widehat{P}(down|D(t)=1))}{100*(1-\widehat{S}_{KM}(t))}}},$$ (2.19)

$$z_{nonevent}(t) = \frac{\widehat{P}(up|D(t)=0) - \widehat{P}(down|D(t)=0)}{\sqrt{\frac{\widehat{P}(up|D(t)=0)(1-\widehat{P}(up|D(t)=0))+\widehat{P}(down|D(t)=0)(1-\widehat{P}(down|D(t)=0))}{100*(\widehat{S}_{KM}(t))}}}.$$ (2.20)

2.2 Time-dependent IDI (tdIDI)

We adopted Chambless and Diao's approach (Chambless and Diao 2006) to generate time-dependent IDI. Again, let $Z = X\widehat{\beta}$ denote the risk score and $D(t)$ denote the corresponding indicator of event by time t. Writing $f(z)$ as the density function of Z, at any given time t and a given cut-off value u $(0 < u < 1)$, we define

$$sensitivity(u,t) = S(u,t) = P(Z>u|D(t)=1) = \int_u^1 f(z|D(t)=1)dz,$$ (2.21)

$$1 - specificity(u,t) = P(u,t) = P(Z>u|D(t)=0) = \int_u^1 f(z|D(t)=0)dz.$$ (2.22)

Using the same set of subscripts, we define tdIDI at time t as

$$tdIDI(t) = (IS_{new}(t) - IS_{old}(t)) - (IP_{new}(t) - IP_{old}(t)),$$ (2.23)

where

$IS_{new}(t) = \int_0^1 sensitivity_{new}(u,t)du, IS_{old}(t) = \int_0^1 sensitivity_{old}(u,t)du, IP_{new}(t) = \int_0^1 (1-specificity_{new}(u,t))du$, and $IP_{old}(t) = \int_0^1 (1 - specificity_{old}(u,t))du.$

From (2.21) and (2.22) and using Bayes' theorem, the conditional probabilities, $f(z|D(t) = 1)$ and $f(z|D(t) = 0)$, can be expressed as

$$f(z|D(t)=1)=\frac{P(D(t)=1|Z=z)f(z)}{P(D(t)=1)}=\frac{(1-S(t|z))f(z)}{\int_z(1-S(t|z))f(z)dz}=\frac{(1-S(t|z))f(z)}{E(1-S(t|Z))} \qquad (2.24)$$

and

$$f(z|D(t)=0)=\frac{P(D(t)=0|Z=z)f(z)}{P(D(t)=0)}=\frac{S(t|z)f(z)}{\int_z S(t|z)f(z)dz}=\frac{S(t|z)f(z)}{E(S(t|Z))} \qquad (2.25)$$

respectively, where $E(\cdot)$ is the expectation function.

With the above expressions, $sensitivity(u,t)$ and $1 - specificity(u,t)$ can be written as

$$\begin{aligned} sensitivity(u,t)&=\int_u^1 f(z|D(t)=1)dz \\ &=\int_u^1 \frac{(1-S(t|z))f(z)}{E(1-S(t|Z))}dz=\frac{1}{E(1-S(t|Z))}\int_u^1(1-S(t|z))f(z)dz, \end{aligned} \qquad (2.26)$$

$$\begin{aligned} 1 - specificity(u,t)&=\int_u^1 f(z|D(t)=0)dz \\ &=\int_u^1 \frac{(S(t|z))f(z)}{E(S(t|Z))}dz=\frac{1}{E(S(t|Z))}\int_u^1(S(t|z))f(z)dz. \end{aligned} \qquad (2.27)$$

Further, the integral of *sensitivity* at time t is defined as $IS(t)=\int_0^1 sensitivity(u,t)du$, which can be expressed as

$$\begin{aligned} IS(t)&=\int_0^1 sensitivity(u,t)du=\int_0^1\int_u^1 f(z|D(t)=1)dzdu \\ &=\frac{1}{E(1-S(t|Z))}\int_0^1\int_u^1(1-S(t|z))f(z)dzdu. \end{aligned} \qquad (2.28)$$

Similarly, the integral of $1 - specificity$ at time t is defined as $IP(t)=\int_0^1(1 - specificity(u,t))du$, and

$$\begin{aligned} IP(t)&=\int_0^1(1 - specificity(u,t)du=\int_0^1\int_u^1 f(z|D(t)=0)dzdu \\ &=\frac{1}{E(S(t|Z))}\int_0^1\int_u^1(S(t|z))f(z)dzdu. \end{aligned} \qquad (2.29)$$

Interchanging the order of integrations within formulas (2.28) and (2.29), we have

$$\begin{aligned} IS(t)&=\frac{1}{E(1-S(t|Z))}\int_0^1\int_0^z(1-S(t|z))f(z)dudz \\ &=\frac{1}{E(1-S(t|Z))}\int_0^1 z(1-S(t|z))f(z)dz \\ &=\frac{E(Z(1-S(t|Z)))}{E(1-S(t|Z))}=\frac{E(Z)-E(ZS(t|Z))}{E(1-S(t|Z))}. \end{aligned} \qquad (2.30)$$

$$\begin{aligned} IP(t)&=\frac{1}{E(S(t|Z))}\int_0^1\int_0^z(S(t|z))f(z)dudz=\frac{1}{E(S(t|Z))}\int_0^1 z(S(t|z))f(z)dz \\ &=\frac{E(ZS(t|Z))}{E(S(t|Z))}. \end{aligned} \qquad (2.31)$$

Substituting the expected values in (2.30) and (2.31) with the corresponding sample means, we can obtain an estimator for $IS(t)$:

$$\widehat{IS}(t) = \frac{\overline{Z} - (\overline{ZS(t|Z)})}{(1 - \overline{S(t|Z)})}. \tag{2.32}$$

Likewise, an estimator for $IP(t)$ is

$$\widehat{IP}(t) = \frac{(\overline{ZS(t|Z)})}{\overline{S(t|Z)}}. \tag{2.33}$$

Thus, the $tdIDI(t)$ defined in (2.23) can be estimated as

$$\widehat{tdIDI}(t) = (\widehat{IS}_{new}(t) - \widehat{IP}_{new}(t)) - (\widehat{IS}_{old}(t) - \widehat{IP}_{old}(t)). \tag{2.34}$$

Note that $IS(t)$ can also be viewed as the weighted average over the range of all cut-off points of *sensitivity* by time t (sample mean of *sensitivity* by time t) and $IS(t)$ can be viewed as the sample mean of $1 - specificity$ over all cut-off points by time t.

The standard deviation of $\widehat{IS}_{new}(t) - \widehat{IP}_{new}(t)$ from equation (2.34), denoted by $\widehat{se}_{new}(t)$, can be calculated as the standard error of paired differences of *sensitivity* and $1 - specificity$ for the new model. Denoting the corresponding estimator for the old model by $\widehat{se}_{old}(t)$ and assuming the correlation coefficient between $\widehat{IS}_{new}(t) - \widehat{IP}_{new}(t)$ for new model and $\widehat{IS}_{old}(t) - \widehat{IP}_{old}(t)$ for old model is γ, we obtain the following asymptotic test for the null hypothesis of $tdIDI(t) = 0$:

$$z(t) = \frac{\widehat{tdIDI}(t)}{\sqrt{(\widehat{se}_{new}(t))^2 + (\widehat{se}_{old}(t))^2 - 2\gamma \widehat{se}_{new}(t)\widehat{se}_{old}(t)}}. \tag{2.35}$$

3. Simulations

Simulation studies were conducted to examine the performance of the proposed tdNRI and tdIDI and compare them to the original NRI and IDI proposed by Pencina et al.

A proportional hazards relationship was generated from a Weibull distribution, which is characterized by two parameters, the scale parameter λ and the shape parameter ν. The corresponding hazard function is given by $h(t|X) = \lambda \nu t^{\nu-1} \exp(\hat{\beta}' X)$, where t represents the time, X the vector of covariates, and $\hat{\beta}$ the vector of estimated regression coefficients. The survival time T of a Cox model with the baseline hazard of a Weibull distribution can be expressed as

$$T = \lambda^{-\frac{1}{\nu}} \left[-\log(U) \times \exp(-\hat{\beta}' X) \right]^{\frac{1}{\nu}} = \left(-\frac{\log(U)}{\lambda \times \exp(-\hat{\beta}X)} \right)^{\frac{1}{\nu}}, \tag{3.1}$$

where $U \sim Uniform(0,1)$.

Two risk variables x_1 and x_2 were simulated with $x_1 \sim N(0, 1)$ and $x_2 \sim Bernoulli(0.5)$, where x_1 was regarded as the baseline risk factor in the original model and x_2 was the additional risk factor of interest in the new model. The estimated coefficient $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)$ was

assumed to be known, where $\hat{\beta}_1 = 3$ and two different values of $\hat{\beta}_2$ were attempted to specify the effect size of the new risk factor x_2 with $\hat{\beta}_2 = 1$ and $\hat{\beta}_2' = 0.5$, respectively. The event time *tevent* was simulated using formula (3.1) with the scale parameter $\lambda = 0.005$ and the shape parameter $\nu = 5$. The censoring time t_{censor} was generated from an exponential distribution $t_{censor} \sim \exp(\lambda')$, and the choice of λ' determines the censoring rate. The observed follow-up time was $t = \min(t_{event}, t_{censor})$.

In this simulation process, data were generated with sample size of 1000 and censoring rates of 15%, 30% and 50% respectively. There were 1000 replicates for each simulation. The simulation results were evaluated in the time period [4, 7] because a large proportion (>70%) of cases occur in this time interval. For comparison, for the same time period, we estimated Pencina's NRI and IDI by assuming that the disease onset status is as at the end of the individual's observation time without considering time-to-event or censoring.

Tables 3.1–3.6 present the simulation results under different censoring mechanisms. Tables 3.1–3.3 contain the simulation results with the hazard ratio (*HR*) of the new risk factor equaling 2.72 ($\hat{\beta}_2 = 1$), and Table 3.4–3.6 contain the results with the *HR* of the new risk factor equaling 1.65 ($\hat{\beta}_2 = 0.5$). The simulation results indicate that the tdNRI is consistently greater than NRI in magnitude under the same conditions, and that the tdNRI and the NRI produce comparable results when the censoring rate is low. As the censoring rate increases, NRI is less sensitive than tdNRI and always overestimates the added discrimination potential of the new risk model. In contrast, tdIDI is consistently smaller than IDI in magnitude, especially when used to assess the discriminatory power of the new marker with small effect size ($\hat{\beta}_2 = 0.5$). Compared with tdIDI, IDI is less sensitive when the censoring rate increases. Not surprisingly, tdNRI and tdIDI perform similarly in detecting the added predictive ability of a new marker with large effect size. However, tdNRI is more sensitive than tdIDI in detecting discriminatory power of the new risk marker with small effect size ($\hat{\beta}_2 = 0.5$).

4. Discussion

We extended the concept of assessing the improvement in model performance by adding new markers from the logistic regression setting with binary outcome to the survival analysis setting with time-to-event outcome, following the definition of NRI and IDI by Pencina et al. Two new statistics, tdNRI and tdIDI were proposed. For each measurement, we derived the sample estimator and asymptotic test. We examined the performance of the tdNRI and tdIDI and compared it with Pencina's NRI and IDI by a series of simulations.

The NRI and IDI are based on binary outcomes in case-control settings. In the context of survival models, where time to event is modeled as a function of baseline exposure factors, NRI or IDI should be estimated as a function of time with proper consideration of censoring. Our proposed tdNRI and tdIDI meet these requirements, thus allowing us to evaluate the improved discriminatory power of a new risk marker in prognostic models for a given time point. As shown by our simulations, compared with Pencina's NRI and IDI that ignore time, our tdNRI and tdIDI both have little bias in detecting the increased power of the model due to the addition of a new marker.

Pencina et al. pointed out in their paper that caution needs to be given to the interpretation of the NRI results due to the property of NRI that it depends on somewhat arbitrary choice of categories, which means NRI can be influenced by the number and extent of the risk categories selected. In our simulations, for clearer illustration, instead of setting up the specific risk categories, we considered the categorization so fine that each person belongs to his/her own risk category. The same is true for the relationship between the NRI and IDI.

The tdIDI is actually the continuous version of the tdNRI, which can explain why the tdNRI and tdIDI perform quite similarly for large samples in our simulations.

We have not addressed the calibration of prognostic survival models thus far. Calibration is another important criterion when evaluating model performance but it may not be very informative in assessing the utility of a new marker. Houwelingen (van Houwelingen 2000) discussed comprehensively how to assess the calibration of the prognostic survival models, which provide us with applicable calibration approaches. Like the original NRI and IDI, tdNRI and tdIDI depend on model calibration as well, so it is important to ascertain that incidence rates in the development and validation sets are similar when assessing the model improvement, especially when the external data are used for model validation. One advantage AUC (or IAUC) possesses over NRI or IDI is that the model calibration issue does not apply to AUC (or IAUC), because they are scale-invariant and independent of model calibration. Thus, as suggested by Pencina et al., the improvement in AUC (or IAUC) should still remain an important criterion when assessing the model improvement.

Because our simple estimator for the tdNRI at time t is given by combining the conditional and unconditional Kaplan-Meier estimators, a potential problem with the tdNRI estimator is that it may be difficult to obtain the conditional Kaplan-Meier estimate when there are very few events/nonevents moving up/down.

Some alternative extensions of the time-dependent IDI may be attempted. One could follow the work by Pepe et al. (Pepe et al. 2008) who showed that the IDIs estimated as differences in discrimination slopes are very close to the logistic R-squares if the model is evaluated on the same sample on which it is developed. The R-squares can be translated to survival analysis settings. Another attempt to obtain the time-dependent IDI could be made by following the bivariate distribution function method to estimate time-dependent sensitivity and specificity proposed by Heagerty et al. (Heagerty et al. 2000). It would be interesting to compare the performance of these two alternative time-dependent IDIs with our proposed tdIDI in future research.

References

Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. Statistics in Medicine. 2006; 25:3474–3486. [PubMed: 16220486]

Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation. 2007; 115:928–935. [PubMed: 17309939]

Eggers KM, Lagerqvist B, Venge P, Wallentin L, Lindahl B. Prognostic Value of Biomarkers During and After Non-ST-Segment Elevation Acute Coronary Syndrome. Journal of the American College of Cardiology. 2009; 54:357–364. [PubMed: 19608034]

Hanley JA, Mcneil BJ. The Meaning and Use of the Area Under A Receiver Operating Characteristic (Roc) Curve. Radiology. 1982; 143:29–36. [PubMed: 7063747]

Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine. 1996; 15:361–387. [PubMed: 8668867]

Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000; 56:337–344. [PubMed: 10877287]

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. Statistics in Medicine. 2008; 27:157–172. [PubMed: 17569110]

Pepe MS, Feng Z, Gu JW. Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al. Statistics in Medicine. 2008; 27:173–181. [PubMed: 17671958]

Ridker PM, Paynter NP, Rifai N, Gaziano M, Cook NR. C-Reactive Protein and Parental History Improve Global Cardiovascular Risk Prediction: The Reynolds Risk Score for Men. Circulation. 2008; 118:S1145.

Shah T, Casas JP, Cooper JA, Tzoulaki I, Sofat R, McCormack V, Smeeth L, Deanfield JE, Lowe GD, Rumley A, Fowkes FGR, Humphries SE, Hingorani AD. Critical appraisal of CRP measurement for the prediction of coronary heart disease events: new data and systematic review of 31 prospective cohorts (vol 38, pg 217, 2009). International Journal of Epidemiology. 2009; 38:890.

Simmons RK, Sharp S, Boekholdt M, Sargeant LA, Khaw KT, Wareham NJ, Griffin SJ. Evaluation of the Framingham risk score in the European Prospective Investigation of Cancer-Norfolk cohort - Does adding glycated hemoglobin improve the prediction of coronary heart disease events? Archives of Internal Medicine. 2008; 168:1209–1216. [PubMed: 18541829]

van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. Statistics in Medicine. 2000; 19:3401–3415. [PubMed: 11122504]

Zethelius B, Berglund L, Sundstrom J, Ingelsson E, Basu S, Larsson A, Venge P, Arnlov J. Use of multiple biomarkers to improve the prediction of death from cardiovascular causes. New England Journal of Medicine. 2008; 358:2107–2116. [PubMed: 18480203]

Table 3.1

The means (mean), standard deviations (s.d.) and the proportions (Prop.) of significant estimators at $\alpha = 0.05$ level from 1000 simulations with a censoring rate of 15% and new risk factor effect with $\hat{\beta}_2 = 1$, evaluated at time points 4, 5, 6, and 7.

Time	tdNRI			NRI			tdIDI			IDI		
	Mean	s.d.	Prop.	Mean	s.d.	Prop.	Mean	s.d.	Prop.	Mean	s.d.	Prop.
4	0.909	0.110	100%	0.401	0.125	86%	0.006	0.002	97%	0.019	0.012	53%
5	0.881	0.140	96%	0.364	0.126	81%	0.005	0.002	88%	0.017	0.011	47%
6	0.887	0.176	76%	0.345	0.120	67%	0.004	0.001	66%	0.016	0.010	47%
7	0.910	0.241	46%	0.327	0.120	44%	0.002	0.001	36%	0.015	0.011	44%

Table 3.2

The means (mean), standard deviations (s.d.) and the proportions (Prop.) of significant estimators at $\alpha = 0.05$ level from 1000 simulations with a censoring rate of 30% and new risk factor effect with $\hat{\beta}_2 = 1$, evaluated at time points 4, 5, 6, and 7.

Time	tdNRI			NRI			tdIDI			IDI		
	Mean	s.d.	Prop.	Mean	s.d.	Prop.	Mean	s.d.	Prop.	Mean	s.d.	Prop.
4	0.912	0.134	100%	0.412	0.089	99%	0.007	0.002	99%	0.017	0.009	65%
5	0.877	0.175	93%	0.374	0.089	98%	0.005	0.002	87%	0.016	0.008	60%
6	0.882	0.240	69%	0.361	0.087	98%	0.004	0.002	60%	0.014	0.009	57%
7	0.895	0.327	46%	0.349	0.087	96%	0.003	0.001	35%	0.014	0.008	57%

Table 3.3

The means (mean), standard deviations (s.d.) and the proportions (Prop.) of significant estimators at $\alpha = 0.05$ level from 1000 simulations with a censoring rate of 50% and new risk factor effect with $\hat{\beta}_2 = 1$, evaluated at time points 4, 5, 6, and 7.

Time	tdNRI			NRI			tdIDI			IDI		
	Mean	s.d.	Prop.	Mean	s.d.	Prop.	Mean	s.d.	Prop.	Mean	s.d.	Prop.
4	0.900	0.161	97%	0.421	0.075	100%	0.006	0.002	97%	0.015	0.008	58%
5	0.865	0.229	87%	0.385	0.077	100%	0.005	0.002	82%	0.013	0.007	52%
6	0.887	0.348	68%	0.378	0.076	100%	0.004	0.002	58%	0.012	0.008	47%
7	0.896	0.523	58%	0.376	0.076	100%	0.003	0.001	37%	0.012	0.008	52%

Table 3.4

The means (mean), standard deviations (s.d.) and the proportions (Prop.) of significant estimators at $\alpha = 0.05$ level from 1000 simulations with a censoring rate of 15% and new risk factor effect with $\hat{\beta}_2 = 0.5$, evaluated at time points 4, 5, 6, and 7.

Time	tdNRI			NRI			tdIDI			IDI		
	Mean	s.d.	Prop.	Mean	s.d.	Prop.	Mean	s.d.	Prop.	Mean	s.d.	Prop.
4	0.779	0.107	98%	0.370	0.119	81%	0.002	0.001	65%	0.007	0.005	34%
5	0.724	0.149	85%	0.331	0.122	76%	0.002	0.001	42%	0.007	0.006	36%
6	0.707	0.190	50%	0.307	0.121	69%	0.001	0.001	21%	0.006	0.005	34%
7	0.685	0.246	20%	0.296	0.118	38%	0.001	0.001	9%	0.006	0.006	34%

Table 3.5

The means (mean), standard deviations (s.d.) and the proportions (Prop.) of significant estimators at $\alpha = 0.05$ level from 1000 simulations with a censoring rate of 30% and new risk factor effect with $\hat{\beta}_2 = 0.5$, evaluated at time points 4, 5, 6, and 7.

Time	tdNRI			NRI			tdIDI			IDI		
	Mean	s.d.	Prop.	Mean	s.d.	Prop.	Mean	s.d.	Prop.	Mean	s.d.	Prop.
4	0.775	0.132	98%	0.379	0.088	100%	0.002	0.001	64%	0.006	0.005	44%
5	0.718	0.178	82%	0.336	0.089	96%	0.001	0.001	39%	0.005	0.004	39%
6	0.699	0.238	49%	0.321	0.087	98%	0.001	0.001	22%	0.005	0.004	35%
7	0.693	0.311	29%	0.315	0.087	98%	0.001	0.001	11%	0.005	0.004	36%

Table 3.6

The means (mean), standard deviations (s.d.) and the proportions (Prop.) of significant estimators at $\alpha = 0.05$ level from 1000 simulations with a censoring rate of 50% and new risk factor effect with $\hat{\beta}_2 = 0.5$, evaluated at time points 4, 5, 6, and 7.

Time	tdNRI			NRI			tdIDI			IDI		
	Mean	s.d.	Prop.	Mean	s.d.	Prop.	Mean	s.d.	Prop.	Mean	s.d.	Prop.
4	0.776	0.182	97%	0.390	0.095	100%	0.002	0.001	58%	0.004	0.004	26%
5	0.722	0.243	74%	0.357	0.087	100%	0.001	0.001	37%	0.005	0.004	26%
6	0.724	0.353	40%	0.343	0.094	100%	0.001	0.001	23%	0.005	0.004	26%
7	0.692	0.542	34%	0.336	0.085	100%	0.001	0.001	12%	0.004	0.004	25%