

Research Article

Evaluating the Incremental Predictive Ability of a New Predictor in Ordinal Outcome Risk Model

Weichieh Wayne Yu¹, Chia-Hao Chang^{2*}, Leng-Chieh Lin³, and Ching-Ho Yen⁴

¹College of Nursing & the Chronic Diseases and Health Promotion Research Center, Chang Gung University of Science and Technology, Taiwan

²Department of Emergency Medicine, Chang Gung Memorial Hospital, Taiwan

³Department of Nursing, Chang Gung University of Science and Technology, Taiwan

⁴Department of Industrial Engineering & Management Information, Huafan University, Taiwan

***Corresponding author**

Chia-Hao Chang, Department of Nursing, Chang Gung University of Science and Technology, Chiayi Campus, Chiayi, Taiwan, 61363, Email: howellchang@gmail.com

Submitted: 26 September 2016

Accepted: 22 March 2017

Published: 24 March 2017

Copyright

© 2017 Chang et al.

OPEN ACCESS**Keywords**

- Model performance
- Integrated discrimination improvement
- Stroke in evolution

Abstract

The integrated discrimination improvement (IDI) index is a popular and sensitive tool in identifying a useful predictor in a binary outcome of interest. However, due to its highly asymmetric null distribution, the Z-statistic proposed by for IDI-based testing of new predictors was proven not valid. In this article, the researchers proposed a modified version of IDI to assess the performance or accuracy of the predictors of an ordinal outcome. Monte Carlo simulations were conducted to investigate the statistical properties of the proposed measure. The asymptotic distribution was derived using U-statistics. The methodology was further applied to a proportional odds ordinal logistic regression model with the inclusion of the blood urea nitrogen/creatinine ratio to discriminate stroke in evolution among acute ischemic stroke patients. These results of the study add to the literature regarding optimal prediction modeling research. The modified version of IDI has a symmetric null distribution; the Z-statistic for testing of new predictors was proven valid.

INTRODUCTION

Using multivariate risk prediction models to identify key predictors associated with the risk levels and quantification of the risk have been among the major advances made in prevention strategies adopted in medicine in the recent decades. In an earlier simulation study of the logistic model [1,2], the estimated average regression coefficients were found to be inaccurate when compared with the “true” population values, even with a fairly large sample. Problems with the validity of the Z-statistic under the null hypothesis were that the Type I error was either greater or less than the nominal value of 5%, and the statistical significance did not imply the improvement in model performance. Equally difficult, was the use of AUC to judge the multitude of increase, because its inability to highlight the value of the new factors that were useful for making predictions [1,3]. To

overcome these problems, the integrated discrimination index (IDI), attempting to assess whether or not adding a new factor improves discrimination performance of an established risk prediction model [4]. It is accomplished by assigning a weight to each movement that is equal to the difference(s) of probabilities based on models with and without a new factor. However, Kerr et al. added that the IDI tends to underestimate the standard error. The proposed Z-statistic in the literature for IDI-based testing of a new predictor was proven to be not valid, because the test statistic did not show a standard normal null distribution [5]. In spite of this, the IDI still has been increasingly popular in predictive modeling research. Furthermore, IDI was developed as binary risk prediction models. In reality, much many predicted outcomes in medicine have been ordinal data [6]. The aim in our real clinical diagnoses was to categorize the National Institutes of Health Stroke Scale (NIHSS) score as ordinal predicted

outcomes. The NIHSS score objectively quantifies the level of impairment caused by a stroke. It has also gained popularity as a clinical tool utilized in treatment planning. For instance, the Bureau of National Health Insurance in Taiwan implemented a payment scale for the rt-PA treatment of acute ischemic stroke (AIS), requiring a minimum of NIHSS score to be equal to or greater than 25 for a patient to receive any form of subsidy. The ordinal outcome models are far more complex to explain the clinical reality than the ones in binary risk prediction. Many valuable clinical data with ordinal responses are often dichotomized for this reason [7-11]. Moreover, assessing the discrimination performance of ordinal risk prediction models is not always a straightforward measure. Most of them are based on the extension of c-statistic, such as generalized c, Obuchowski, average dichotomized c and ordinal volume under the ROC surface [12-17]. However, none of which can be used to test for the incremental value of a predictor. In light of this, the researchers proposed a new measure to evaluate the incremental gain of a new predictor in an ordinal risk prediction model. Ideally, a fine discrimination performance measure should possess the following two properties: P1. The power of the measure should be approximately equal to the stated significance level when H_0 is true, P2. The null distribution of the measure follows a standard normal distribution. As previously mentioned, statistical IDI does not satisfy both assumptions. This provides a ground for improvements.

PROPOSED METHODOLOGY

Modified version of IDI for ordinal outcomes

Assuming $p_{i1}, p_{i2}, \dots, p_{in_i}$, in which $i = 1, 2, 3$ represent the independently randomly predicted probabilities of the event from three overlapping and continuous distributions. The volume under the ROC surface (VUS) is equal to the probability that three predicted probabilities, one from each category, are classified in the correct order category [16]. Hence, the unbiased nonparametric estimator of the VUS proposed by Dreiseitl et al. is equivalent to the nonparametric test statistic proposed by Terpstra and Magel except that it multiplies the weight $\frac{1}{n_1 n_2 n_3}$ by the estimator, which is expressed as follows [17,18]:

$$VUS = \frac{1}{n_1 n_2 n_3} \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \sum_{j_3=1}^{n_3} I(p_{1j_1} < p_{2j_2} < p_{3j_3}) \quad (1)$$

Specifically, $I(p_{1j_1} < p_{2j_2} < p_{3j_3}) = 1$ provided that at least one strict inequality exists; otherwise,

$$I(p_{1j_1} < p_{2j_2} < p_{3j_3}) = 0.$$

Next, the following notations are introduced:

$F(k)_{\text{new}}$ is the fraction of the rank of p_k being equal to k with respect to p_1, p_2, \dots, p_k (numerator) over $\prod_{i=1}^k n_i$ (denominator), and the subscript new refers to the model with a new factor.

$F(k)_{\text{old}}$ is the fraction of the rank of p_k being equal to k with respect to p_1, p_2, \dots, p_k (numerator) over $\prod_{i=1}^k n_i$ (denominator), and the subscript old refers to the model without a new factor. Similarly, $F(i)_{\text{new}}$ and $F(i)_{\text{old}}$, where $i = 1, \dots, k-1$, correspond to the fractions of rank of p_i being equal to i with respect to p_1, p_2, \dots, p_k (numerator) among $\prod_{i=1}^k n_i$ (denominator), only that the subscript new refers to the model with a new factor and subscript old refers to the model without one.

In a set of cases, it is desirable that the predicted probabilities of the event of each category are consistent with the order of the ordinal outcome. Therefore, the score awarded to a set equals the number of categories for which this holds. The new measure of three ordinal categories is expressed as the following form $\text{MIDI}(3) = [F(3)_{\text{new}} - F(3)_{\text{old}}] + [F(2)_{\text{new}} - F(2)_{\text{old}}] + [F(1)_{\text{new}} - F(1)_{\text{old}}]$

$$= [F(3)_{\text{new}} + F(2)_{\text{new}} + F(1)_{\text{new}}] - [F(3)_{\text{old}} + F(2)_{\text{old}} + F(1)_{\text{old}}]$$

$$= SF(3)_{\text{new}} - SF(3)_{\text{old}}$$

$\text{MIDI}(3)$ is denoted as the modified version of IDI of three ordinal categories and $SF(3)$ refers to the sum of $F(3)$, $F(2)$ and $F(1)$. Based on the equation (1), the $\text{MIDI}(3)$ can be inferred as follows:

$$\begin{aligned} \text{MIDI}(3) &= SF(3)_{\text{new}} - SF(3)_{\text{old}} \\ &= \frac{1}{n_1 n_2 n_3} \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \sum_{j_3=1}^{n_3} [I(R(p_{1j_1}) = 1) + I(R(p_{2j_2}) = 2) + I(R(p_{3j_3}) = 3)] \\ &\quad - \frac{1}{n_1 n_2 n_3} \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \sum_{j_3=1}^{n_3} [I(R(p_{1j_1}) = 1) + I(R(p_{2j_2}) = 2) + I(R(p_{3j_3}) = 3)] \end{aligned}$$

$R(p_i)$ is denoted the rank of p_i with respect to p_1, p_2 , and p_3 , and $I(\cdot)$ denotes the indicator function.

The following was obtained as a way of generalization:

$$\begin{aligned} \text{MIDI}(k) &= \text{SF}(k)_{\text{new}} - \text{SF}(k)_{\text{old}} \\ &= \frac{1}{n_1 n_2 \cdots n_k} \sum_{j_1=1}^{n_1} \cdots \sum_{j_k=1}^{n_k} [I(R(p_{1j_1})=1) + \cdots + I(R(p_{kj_k})=k)] \\ &\quad - \frac{1}{n_1 n_2 \cdots n_k} \sum_{j_1=1}^{n_1} \cdots \sum_{j_k=1}^{n_k} [I(R(p_{1j_1})=1) + \cdots + I(R(p_{kj_k})=k)] \end{aligned}$$

Consequently, also obtained was an asymptotic test for null hypothesis of $\text{MIDI}(k) = 0$ as follows:

$$Z_{\text{MIDI}(k)} = \frac{\text{MIDI}(k)}{\sqrt{\sigma_{\text{SF}(k),\text{new}}^2 + \sigma_{\text{SF}(k),\text{old}}^2 - 2\hat{\rho}_{\text{new,old}}\sigma_{\text{SF}(k),\text{new}}\sigma_{\text{SF}(k),\text{old}}}}, \quad (2)$$

In the equation (2), $\hat{\rho}_{\text{new,old}}$ is the estimator of correlation coefficient of $\text{SF}(k)_{\text{new}}$ and $\text{SF}(k)_{\text{old}}$ under H_0 .

$$\sigma_{\text{SF}(k)}^2 = \text{Var}_0[\text{SF}(k)] = v_0^2 + \sum_{i=1}^{k-1} v_i^2,$$

where $v_0^2 = \frac{1 - (1/k)}{n_1 \cdots n_k}$ represents no tie. In the case of

i ties, [19] Presented an algorithm for the computation of

$$\text{the } \sum_{i=1}^{k-1} v_i^2.$$

The sampling distribution of MIDI and Z_{MIDI}

In this section, simulation theory was used to attest the accuracy of the standard deviation estimate used in (2), the normality of the sample distribution of $\text{MIDI}(k)$, and null distribution of $Z_{\text{MIDI}(k)}$ for $k = 3$. Assuming that the predicted probabilities of the triggering event (level 3 for $k=3$ and level 4 for $k=4$) were generated under the ordinal logistic regression model: Ordered logit $p(s = k | x) = \text{logit}(p_{\text{old}}) = \beta_0 + \beta_1 x$, where s denotes the symptom level and x is the established predictor, which were assumed from iid $N(1, 1)$. Moreover, s replicates the sequence of 1 to 3 or 4 n_1 , n_2 and n_3 or n_4 times, respectively. A new model with a candidate predictor y is also considered; that is, Ordered logit $p(s=k | x, y) = \text{logit}(p_{\text{new}}) = \beta_0 + \beta_1 x + \beta_2 y$, where y are simulated based on the coefficient β_2 of 0, 2, 4 and 6.

The researchers assumed the above-mentioned conditions using the software R 3.0.1 (R Development Core Team, 2013) and equip a proportional odds logistic regression model to them using the MASS and rms package in order to obtain predicted probabilities of the event and p-value for β_2 coefficient in the fitted ordinal logistic regression model. This Monte Carlo simulation was replicated 1,000 times

at $\alpha = 0.05$ using different combinations of n_i . Taking into account of the simulated data sets with an established predictor x and a candidate predictor of symptom level that has no predictive capacity, the proportional odds logistic simulation model was used and coefficient β_2 was set to be equal to 0. The accuracy of the standard error estimate was first investigated:

$$\sqrt{2\sigma_{\text{MIDI}(k)}^2 - 2\hat{\rho}_{\text{new,old}}\sigma_{\text{MIDI}(k)}^2} \text{ used in (2). Secondly, the}$$

researchers simulated 1,000 data sets to obtain $\hat{\rho}_{\text{new,old}}$ for different sample size combination and therefore computed $Z_{\text{MIDI}(k)}$. The simulated powers of $Z_{\text{MIDI}(3)}$ are shown to be close to the nominal level under the null hypothesis. (Table 1) illustrates the simulated critical values for $\alpha = 0.05$ of sample size combinations. Comparing with asymptotic critical values of 1.6449 for $\alpha = 0.05$ under the null hypothesis, they are all not too far apart. We also investigated more thoroughly of the sampling distribution of MIDI and Z_{MIDI} under H_0 . (Figures 1,2) show the histograms of $\text{MIDI}(3)$, $Z_{\text{MIDI}(3)}$ and $\text{MIDI}(4)$, $Z_{\text{MIDI}(4)}$ under H_0 of the sample size combinations (10, 10, 10), (150, 10, 50) and (10, 10, 10, 10), (150, 10, 10, 10), and $M = 10,000$. The null distributions of MIDI and Z_{MIDI} are all symmetric with are centered in zero. (Table 2) provides the p values of the 5 randomly individual cases of β_2 coefficients of 2, 4 and 6 in the fitted ordinal logistic regression model and for the $\text{MIDI}(3)$ and $\text{MIDI}(4)$ with sample size combinations (20, 20, 20) and (20, 20, 20, 20). The result demonstrated that the p-values of β_2 are inconsistent with the p-values of Z_{MIDI} . The resulting p-values of β_2 are only significant for a correlation between s and y . In contrast to this, the MIDI is used to assess improvement in model discrimination. However, small p values of β_2 cannot guarantee to detect the significance of MIDI (Table 1,2).

CASE STUDY

Up to 40% of the patient with the stroke experienced early deterioration in their conditions after being hospitalized [20]; such patients are defined as stroke in evolution (SIE). SIE patients experience a worsening neurologic condition as indicated by an increase in their NIHSS scores within 72 hours of admission. Theoretically, although any increase in NIHSS score can be indicative of SIE, symptoms can be border line in some cases. Hence, any change of below 0 is categorized as amelioration, 1-2 as border line, and 3 or more as SIE in the present study. The reason for such is that a larger increase can often be more indicative of the patient's worsening conditions. The analysis was based on data from 196 AIS patients recruited from the Chia-Yi Chang Gung Memorial Hospital and 27 of which were diagnosed of SIE.

Table 1: Null distribution of $Z_{MIDI}(k)$ for $k = 3$.

| k = 3 | | | | | | | | | |
|-------|-----------------|--|--|------------------------|----|-----------------|--|--|------------------------|
| | Average MIDI(3) | Average power $P(Z_{MIDI}(3) > 1.6449)$ | Simulated critical value $Z_{MIDI}(3)_{0.05}$ | $\hat{\rho}_{new,old}$ | | Average MIDI(3) | Average power $P(Z_{MIDI}(3) > 1.6449)$ | Simulated critical value $Z_{MIDI}(3)_{0.05}$ | $\hat{\rho}_{new,old}$ |
| 1 | 0.00081 | 0.0533 | 1.6982 | 0.9644 | 20 | 0.00111 | 0.0387 | 1.4321 | 0.9527 |
| 2 | -0.0002 | 0.0444 | 1.5149 | 0.8736 | 21 | -0.0020 | 0.0377 | 1.3672 | 0.9182 |
| 3 | 0.00946 | 0.0567 | 1.7640 | 0.8505 | 22 | 0.00155 | 0.0441 | 1.5611 | 0.9614 |
| 4 | 0.00355 | 0.0531 | 1.6965 | 0.8442 | 23 | -0.0005 | 0.0354 | 1.3421 | 0.9266 |
| 5 | 0.00478 | 0.052 | 1.6884 | 0.8185 | 24 | -0.0012 | 0.0374 | 1.3813 | 0.9305 |
| 6 | 0.00000 | 0.0473 | 1.5961 | 0.9810 | 25 | 0.00113 | 0.0385 | 1.4261 | 0.9529 |
| 7 | 0.00442 | 0.0473 | 1.5834 | 0.9113 | 26 | 0.00179 | 0.0529 | 1.6852 | 0.9727 |
| 8 | 0.00007 | 0.0413 | 1.4856 | 0.8867 | 27 | 0.00132 | 0.0475 | 1.5985 | 0.9534 |
| 9 | 0.00099 | 0.0424 | 1.5240 | 0.8713 | 28 | -0.0001 | 0.0395 | 1.4346 | 0.9183 |
| 10 | 0.00426 | 0.0513 | 1.6645 | 0.8439 | 29 | -0.0026 | 0.0404 | 1.4417 | 0.8812 |
| 11 | 0.00044 | 0.0489 | 1.6251 | 0.9864 | 30 | 0.00333 | 0.0447 | 1.5636 | 0.9368 |
| 12 | -0.0036 | 0.0326 | 1.2634 | 0.9344 | 31 | 0.00006 | 0.0460 | 1.6012 | 0.9962 |
| 13 | 0.00024 | 0.0401 | 1.4233 | 0.9050 | 32 | 0.00023 | 0.0413 | 1.4856 | 0.9913 |
| 14 | -0.0006 | 0.0409 | 1.4667 | 0.8860 | 33 | -0.0021 | 0.0406 | 1.4593 | 0.9403 |
| 15 | -0.0070 | 0.0349 | 1.3398 | 0.8695 | 34 | 0.00159 | 0.0368 | 1.2572 | 0.9758 |
| 16 | -0.0009 | 0.0363 | 1.3824 | 0.9605 | 35 | 0.00016 | 0.0529 | 1.6852 | 0.9610 |
| 17 | 0.00082 | 0.0406 | 1.4593 | 0.9206 | 36 | 0.00085 | 0.0465 | 1.4635 | 0.9419 |
| 18 | 0.00088 | 0.0466 | 1.5920 | 0.9739 | 37 | 0.00408 | 0.0452 | 1.4772 | 0.9168 |
| 19 | 0.00115 | 0.0407 | 1.4558 | 0.9420 | 38 | 0.00120 | 0.0475 | 1.6116 | 0.8540 |

k = 3:
1: $n_1=10, n_2=10, n_3=10$; 2: $n_1=30, n_2=10, n_3=10$; 3: $n_1=50, n_2=10, n_3=10$; 4: $n_1=70, n_2=10, n_3=10$;
5: $n_1=150, n_2=10, n_3=10$; 6: $n_1=10, n_2=30, n_3=10$; 7: $n_1=30, n_2=30, n_3=10$; 8: $n_1=50, n_2=30, n_3=10$;
9: $n_1=70, n_2=30, n_3=10$; 10: $n_1=150, n_2=30, n_3=10$; 11: $n_1=10, n_2=50, n_3=10$; 12: $n_1=30, n_2=50, n_3=10$;
13: $n_1=50, n_2=50, n_3=10$; 14: $n_1=70, n_2=50, n_3=10$; 15: $n_1=150, n_2=50, n_3=10$; 16: $n_1=24, n_2=10, n_3=20$;
17: $n_1=24, n_2=10, n_3=47$; 18: $n_1=24, n_2=10, n_3=23$; 19: $n_1=24, n_2=10, n_3=17$; 20: $n_1=24, n_2=10, n_3=31$;
21: $n_1=22, n_2=10, n_3=44$; 22: $n_1=22, n_2=10, n_3=19$; 23: $n_1=22, n_2=10, n_3=37$; 24: $n_1=22, n_2=10, n_3=38$;
25: $n_1=22, n_2=10, n_3=29$; 26: $n_1=16, n_2=10, n_3=16$; 27: $n_1=16, n_2=10, n_3=13$; 28: $n_1=16, n_2=10, n_3=29$;
29: $n_1=16, n_2=10, n_3=49$; 30: $n_1=16, n_2=10, n_3=23$; 31: $n_1=82, n_2=79, n_3=83$; 32: $n_1=82, n_2=79, n_3=87$;
33: $n_1=65, n_2=143, n_3=24$; 34: $n_1=70, n_2=92, n_3=58$; 35: $n_1=82, n_2=56, n_3=59$; 36: $n_1=67, n_2=54, n_3=33$;
37: $n_1=62, n_2=67, n_3=25$; 38: $n_1=154, n_2=16, n_3=27$

According to Lin, the ratio of blood urea nitrogen/creatinine (BUN/Cr) higher than 15 was cogent predictor of SIE [21]. We evaluated the improvement in model performance with the inclusion of age and BUN/Cr ratio. $Z_{MIDI(3)}$ was used to test the significance. The increase value in the VUS was also presence for reference. Thirty (15.3%) of 196 patients were diagnosed with SIE. The mean (SD) age of the 196 participants enrolled was 70.6 (10.4) years, ranging from 44 to 99 years. A proportional odds ordinal logistic regression model showed that the BUN/Cr ratio higher than 15 was significant ($p = 0.011$), and VUS increased from 0.1943 to 0.3212 adding that BUN/Cr ratio is higher than 15 (Table 3). The estimator of correlation coefficient of $SF(3)_{new}$ and $SF(3)_{old}$ under H_0 of 0.854 was used based on the $\hat{\rho}_{new,old}$ for sample

size scenario 38 (Table 3). The values of F (3), F (2) and F (1) with BUN/Cr are shown the top row in Table (4). The significant difference was noted in $Z_{MIDI(3)}$ for BUN/Cr. MIDI(3), with the inclusion of BUN/Cr, was summed up to 0.3681, due to a 2.82 % increase in F (1), 12.08 % increase in F (2) and 23.55 % increase in F (3), respectively (Table 3,4).

The result is supported by the findings presented by Lin et al., indicating that BUN/Cr was found to be the significant predictor of SIE [22]. In addition, MIDI (3) not only suggested that the inclusion of the factor in the proportional odds ordinal logistic regression model can result in improvement in performance, but also provided the F (.) values of for three-level outcome from this study. We may conclude that the addition of BUN/Cr improved the F (.) values by 2.81 %,

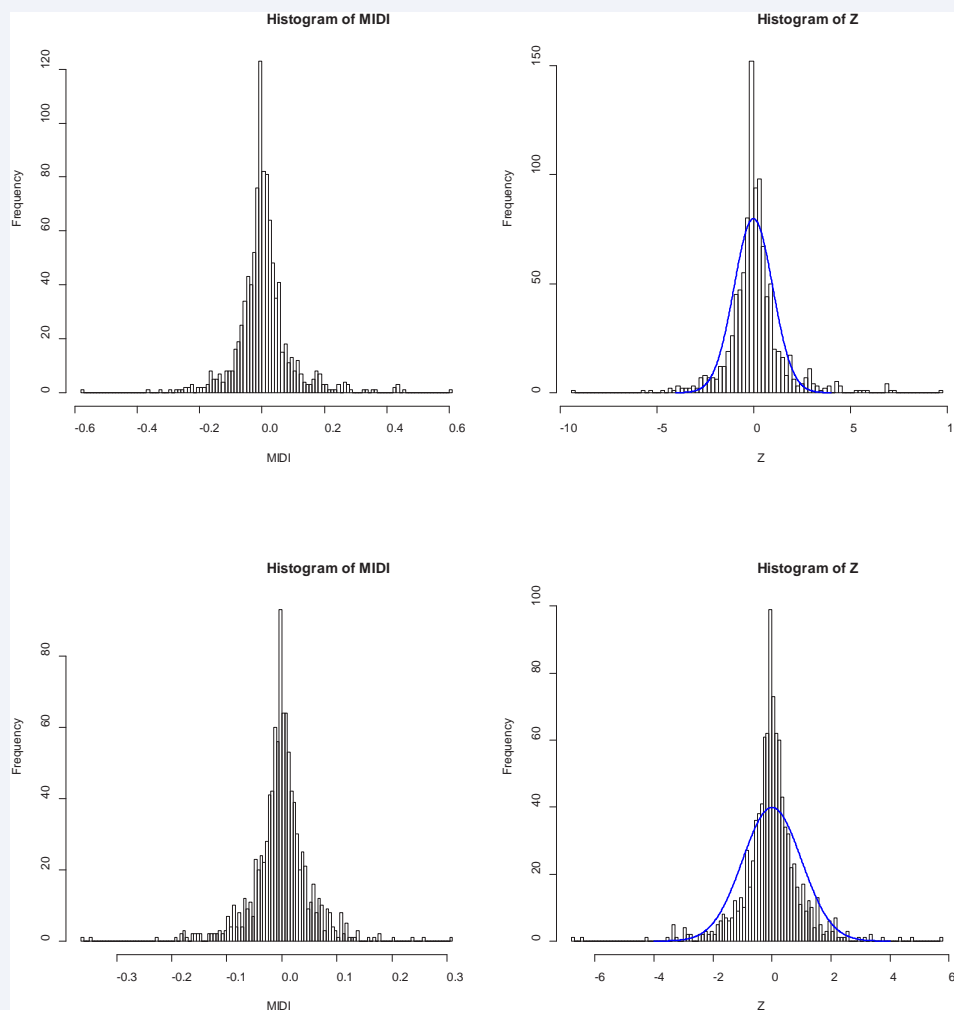


Figure 1 Null distribution of $MIDI(3)/Z_{MIDI(3)}$ for simulated data sets using a proportional odds logistic model & Standard normal curves of $Z_{MIDI(3)}$ (top row; $n_1 = n_2 = n_3 = 10$, bottom row; $n_1 = 150$, $n_2 = 10$, $n_3 = 50$).

Table 2: P-values of the null hypothesis in the proportional odds logistic simulation model and corresponding MIDI (k) for $k=3$ and 4.

| k=3 | True β_2 | p value of β_2 | True MIDI(3) | p value of MIDI(3) | k=4 | True β_2 | p value of β_2 | True MIDI(3) | p value of MIDI(3) |
|-------------|----------------|----------------------|--------------|--------------------|-------------|----------------|----------------------|--------------|--------------------|
| $\beta_2=2$ | | | | | $\beta_2=2$ | | | | |
| 1 | 2.1063 | 0.257 | 0.381 | <0.0001 | 1 | 2.1063 | 0.257 | 0.381 | <0.0001 |
| 2 | 2.1423 | 0.264 | 0.144 | 0.0102 | 2 | 2.1423 | 0.264 | 0.144 | 0.0102 |
| 3 | 2.132 | 0.285 | 0.083 | 0.0918 | 3 | 2.132 | 0.285 | 0.083 | 0.0918 |
| 4 | 1.2444 | 0.513 | -0.004 | 0.5255 | 4 | 1.2444 | 0.513 | -0.004 | 0.5255 |
| 5 | 1.963 | 0.297 | -0.434 | 1 | 5 | 1.963 | 0.297 | -0.434 | 1 |
| $\beta_2=4$ | | | | | $\beta_2=4$ | | | | |
| 1 | 4.57 | 0.05 | -0.052 | 0.7976 | 1 | 4.57 | 0.05 | -0.052 | 0.7976 |
| 2 | 4.2229 | 0.026 | -0.103 | 0.9505 | 2 | 4.2229 | 0.026 | -0.103 | 0.9505 |
| 3 | 3.5462 | 0.046 | 0.233 | <0.0001 | 3 | 3.5462 | 0.046 | 0.233 | <0.0001 |
| 4 | 4.2862 | 0.029 | 0.116 | 0.0316 | 4 | 4.2862 | 0.029 | 0.116 | 0.0316 |
| 5 | 4.5433 | 0.028 | 0.356 | <0.0001 | 5 | 4.5433 | 0.028 | 0.356 | <0.0001 |
| $\beta_2=6$ | | | | | $\beta_2=6$ | | | | |
| 1 | 8.2746 | 6E-04 | -0.104 | 0.9522 | 1 | 8.2746 | 6E-04 | -0.104 | 0.9522 |
| 2 | 6.4363 | 0.003 | 0.285 | <0.0001 | 2 | 6.4363 | 0.003 | 0.285 | <0.0001 |
| 3 | 7.0936 | 0.002 | 0.045 | 0.2355 | 3 | 7.0936 | 0.002 | 0.045 | 0.2355 |
| 4 | 7.4539 | 0.004 | -0.132 | 0.9828 | 4 | 7.4539 | 0.004 | -0.132 | 0.9828 |
| 5 | 6.5886 | 0.002 | 0.231 | 0.0001 | 5 | 6.5886 | 0.002 | 0.231 | 0.0001 |

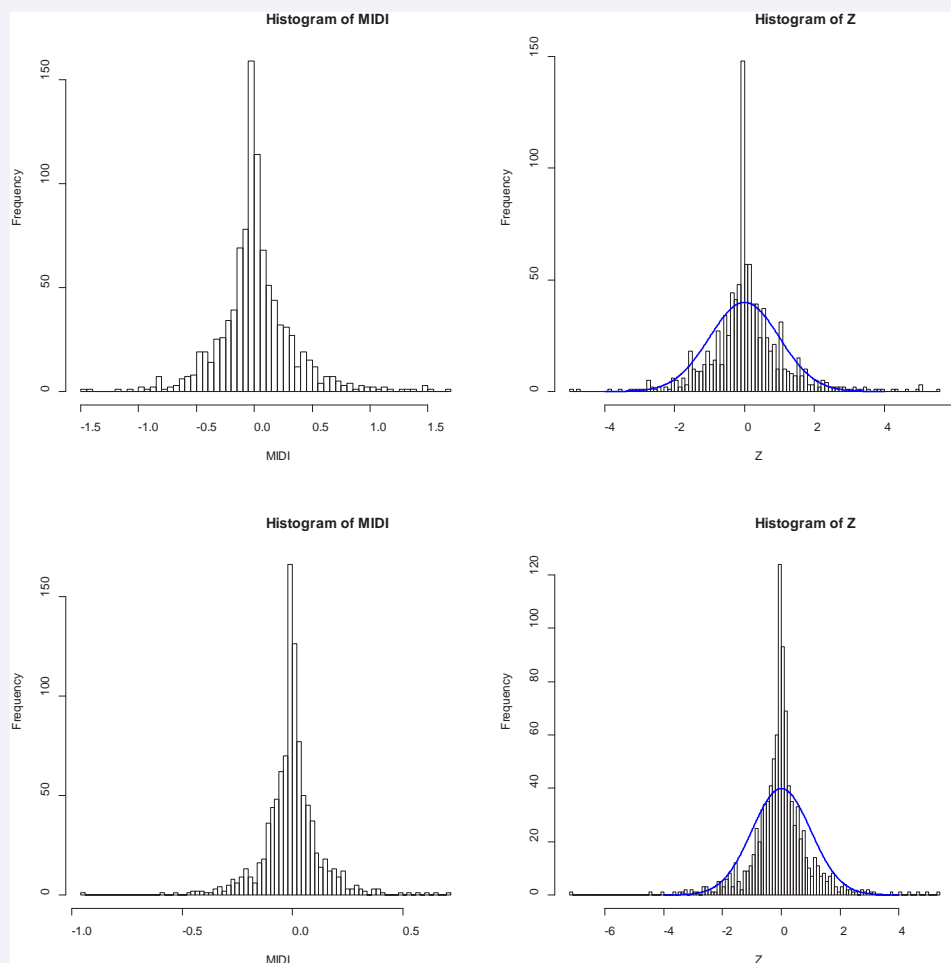


Figure 2 Null distribution of MIDI(4)/ZMIDI(4) for simulated data sets using a proportional odds logistic model & Standard normal curves of ZMIDI(4) (top row; $n_1 = n_2 = n_3 = n_4 = 10$, bottom row; $n_1 = 150, n_2 = 10, n_3 = 10, n_4 = 10$).

Table 3: Proportional odds ordinal logistic regression coefficients for examined predictors.

| Variable | Estimate | Std. Error | p value | 95% Confidence Interval | |
|-------------------------------|----------|------------|---------|-------------------------|---------|
| Age | 0.031 | 0.018 | 0.086 | -0.004 | 0.066 |
| Initial NIHSS score ≥ 12 | -0.555 | 0.378 | 0.142 | -1.296 | 0.185 |
| Age | 0.032 | 0.018 | 0.065 | -0.002 | 0.067 |
| Glasgow Coma Scale ≤ 12 | 0.779 | 0.399 | 0.051 | -0.003 | 1.560 |
| Age | 0.033 | 0.019 | 0.077 | -0.004 | 0.070 |
| D-dimers > 1000 | -0.119 | 0.378 | 0.753 | -0.859 | 0.621 |
| Age | 0.037 | 0.018 | 0.036 | 0.002 | 0.071 |
| BUN/Cr > 15 | -0.900 | 0.356 | 0.011 | -1.598 | -0.202 |
| Age | 0.0353 | 0.0189 | 0.0623 | -0.0018 | 0.0724 |
| Initial NIHSS score ≥ 12 | -0.1963 | 0.4961 | 0.6923 | -1.1686 | 0.7760 |
| Glasgow Coma Scale ≤ 12 | 0.5403 | 0.5175 | 0.2964 | -0.4739 | 1.5546 |
| D-dimers > 1000 | 0.1127 | 0.4140 | 0.7854 | -0.6986 | 0.9241 |
| BUN/Cr > 15 | -0.7594 | 0.3735 | 0.0420 | -1.4915 | -0.0274 |

12.08 % and 23.55 % for cases with NIHSS score increases of below 0, 1–2 and 3 or more among $\prod_{i=1}^3 n_i$ results. We also evaluated the improvement in model performance with the inclusion of NIHSS, Glasgow Coma Scale score

(GCS) and D-dimers using the MIDI (3). With the case of BUN/Cr, F (2) (12.08%) and F(3) (23.55 %) made most contributions to MIDI(3), and F(1) (2.81%) is the smallest among NIHSS, GCS and D-dimers. This implied that the discrimination performance of BUN/Cr is not better than

Table 4: Contributions to SIE prediction in AIS patients¹.

| variable | VUS | F(1) | F(2) | F(3) | MIDI(3) | p value |
|---|--------|--------|--------|--------|---------|---------|
| Effect of adding variables to model with age only | | | | | | |
| With age only | 0.1943 | 0.4382 | 0.2938 | 0.3388 | | |
| + BUN/Cr | 0.3212 | 0.4501 | 0.4146 | 0.5743 | 0.3212 | 2.9e-10 |
| + NIHSS | 0.2421 | 0.4610 | 0.3422 | 0.4019 | 0.1343 | 0.0120 |
| + GCS | 0.2758 | 0.4602 | 0.3776 | 0.4688 | 0.2358 | 3.6e-05 |
| + D-dimers | 0.2169 | 0.4373 | 0.3133 | 0.3638 | 0.0436 | 0.2318 |
| Effect of deleting variables from full model ² | | | | | | |
| Full model | 0.3761 | 0.4659 | 0.4430 | 0.6700 | | |
| - BUN/Cr | 0.3105 | 0.4754 | 0.4100 | 0.5265 | 0.1670 | 0.002 |
| - NIHSS | 0.3377 | 0.4563 | 0.4153 | 0.6212 | 0.0861 | 0.074 |
| - GCS | 0.3367 | 0.4719 | 0.4293 | 0.5817 | 0.0959 | 0.053 |
| - D-dimers | 0.3558 | 0.4632 | 0.4373 | 0.6294 | 0.0489 | 0.205 |

+ indicates the addition of each predictor separately to the model with age only; -, the deletion of each predictor separately from the full model.
¹Estimated from Proportional odds ordinal logistic regression model.
²The model with age, initial NIHSS, GCS, D-dimers and BUN/Cr.

NIHSS, GCS and D-dimers for the predicted probability of SIE for patients without early deterioration after stroke. Moreover, the bottom row in (Table 2) also shows the 1st step of backward elimination from full model (including age, BUN/Cr, NIHSS, GCS and D-dimers). The NIHSS and GCS were far from significant in proportional odds ordinal logistic regression model with $p = 0.6923$ and 0.2964 , respectively. However, the MIDI (3) identified them as marginal significant predictors ($p = 0.074$, $p = 0.053$). With the addition of BUN/Cr, F (.) values improved by 14.34 % and 3.30 % for cases with NIHSS increases of 3 or more

and 1-2 among $\prod_{i=1}^3 n_i$ results, but a loss of 0.95 % was

noted for cases whose NIHSS increases were below 0. Subsequently, the addition of NIHSS improved F(.) values by 4.88 %, 2.77 % and 0.96 %, and the addition of GCS improved F(.) values by 8.82 %, 1.37 % and -0.60 %. The F (1) values increased quite evenly throughout from 0.4659 to 0.4754 and 0.4719, as BUN/Cr and GCS were removed separately from the full mode.

DISCUSSION

The example in Table (2) provided in the study also illustrates how an increase in the MIDI (3) can lead to a decrease in the F (1). This might suggest the researchers to reconsider alternative ways to evaluate increases in the MIDI (3) or VUS, or perhaps the p value of <0.05 shall not be suggested in the clinical practice. In clinical reality, most patients' symptoms may not be considered critical by medical professionals. However, the additional cost of obtaining an extra predictor is significantly high. Thus, it might not be worth the efforts to include it in the

prediction models. Alternatively, we might want to resort to traditionally measures. A word of caution as always, any decision of the kind should be based upon the clinical and public health implications.

Finally, it should be noted that one of the limitations of the proposed measure is that to estimate the unknown correlation coefficient between old and new SF under H_0 , $\hat{\rho}_{new,old}$, may result in intensive computation. Note that the data given in (Table 3) can be used to approximate $\hat{\rho}_{new,old}$ for sample size combinations that do not appear in the table. For instance, one might consider a parametric or nonparametric fit to the data $\{n_1, n_2, n_3, \rho_{new,old}\}$ for $k=3$ and then applies this fit to obtain $\hat{\rho}_{new,old}$.

REFERENCES

1. Pencina MJ. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med. 2011; 30: 11-21.
2. Lilienfeld DE, Pyne DA. The logistic analysis of epidemiologic prospective studies: investigation by simulation. Stat Med. 1984; 3: 15-26.
3. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation. 115; 928-935.
4. Pencina MJ. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. Stat Med. 2008; 27: 157-172.
5. Kerr KF, Mc Clelland RL, Brown ER, Lumley T. Evaluating the incremental value of new biomarkers with integrated discrimination improvement. Am J Epidemiol. 2011; 174: 364-374.
6. Whitehead A, Omar RZ, Higgins JP, Savalun E, Turner RM, Thompson SG, et al. Meta-analysis of ordinal outcomes using individual patient data. Stat Med. 2001; 20: 2243-2260.
7. Hacke W, Kaste M, Bluhmki E, Brozman M, Dávalos A, Guidetti D, et

- al. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. *N Engl J Med*. 2008; 359: 1317-1329.
8. Chen MY, Wang EK, Jeng YJ. Adequate sleep among adolescents is positively associated with health status and health-related behaviors. *BMC Public Health*. 2006; 6: 59.
9. Yang RJ, Wang EK, Hsieh YS, Chen MY. Irregular breakfast eating and health status among adolescents in Taiwan. *BMC Public Health*. 2006; 6: 295.
10. De Lemos JA, Drazner MH, Omland T, Ayers CR, Khera A, Rohatgi A, et al. Association of troponin T detected with a highly sensitive assay and cardiac structure and mortality risk in the general population. *JAMA*. 2010; 304: 2503-2512.
11. Guo SE, Huang TJ, Huang JC, Lin MS, Hong RM, Chang CH, et al. Alcohol, betel-nut and cigarette consumption are negatively associated with health promoting behaviors in Taiwan: A cross-sectional study. *BMC Public Health*. 2013; 13: 257.
12. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982; 247: 2543-2546.
13. Harrell FEJ. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Springer. 2001.
14. Obuchowski NA. Estimating and comparing diagnostic tests' accuracy when the gold standard is not binary. *Acad Radiol*. 2005; 12: 1198-1204.
15. Waegeman W, De Baets B, Boullart L. ROC Analysis in Ordinal Regression Learning, *Pattern Recognition Letters*. 2008; 29: 1-9.
16. Scurfield BK. Multiple-event forced-choice tasks in the theory of signal detectability, *J Math Psychol*. 1996; 40: 253-269.
17. Mossman D. Three-way ROCs. *Med Decis Making*. 1999; 19: 78-89.
18. Dreiseitl S, Ohno-Machado L, Binder M, Comparing three-class diagnostic tests by three-way ROC analysis, *Medical Decision Making*. 2000; 20: 323-331.
19. Chang CH, Chin CC, Yu W, Huang YY. Using the Bernoulli trial approaches for detecting ordered alternatives. *BMC Medical Research Methodology*. 2013; 13: 148-154.
20. Terpstra JT, Magel RC. A new nonparametric test for the ordered alternative problem, *Nonparametric Statistics*. 2003; 15: 289-301.
21. Davalos A, Castillo J. Progressing stroke. In: Fisher M, Bogousslavsky J. Current review of cerebrovascular disease, Philadelphia: Current Medicine Inc. 1999; 149-160.
22. Lin LC, Yang JT, Weng HH, Hsiao CT, Lai SL, Fann WC, et al. Predictors of early clinical deterioration after acute ischemic stroke. *Am J Emerg Med*. 2011; 29: 577-581.

Cite this article

Yu WW, Chang CH, Lin LC, Yen CH (2017) Evaluating the Incremental Predictive Ability of a New Predictor in Ordinal Outcome Risk Model. *JSM Health Educ Prim Health Care* 2(1): 1026.