

## 两种比较模型预测效果的评价指标及其 R 实现\*

李彤阳<sup>#</sup>, 林卓琛<sup>#</sup>, 葛琪, 陈雯, 张晋昕<sup>△</sup>

510080 广州, 中山大学 公共卫生学院

**[摘要]** 目的: 介绍净重新分类指数与整体鉴别指数, 阐明两个统计指标在医学研究中普遍使用的 Logistic 回归模型预测效果评价中的意义与实现。方法: 给出净重新分类指数与整体鉴别指数的定义式, 结合实例说明在 R 中的计算实现, 并与常用的预测效果评价指标: 受试者工作特征曲线, 曲线下面积, 以及灵敏度、特异度等进行对比, 阐述其应用方面的特点。结果: 在 Logistic 回归模型应用中, 净重新分类指数与整体鉴别指数可以量化出模型筛选中新模型的改进效果, 在 R 软件中可以方便地获得此效果的估计值及假设检验结果。结论: 净重新分类指数与整体鉴别指数定量地反映了模型优化前后的差异, 具有实际运用价值。运用 R 软件可完成相关计算。

**[关键词]** 预测模型; 评价指标; 净重新分类指数; 整体鉴别指数

**[中图分类号]** R181.2+3; R311 **[文献标志码]** A doi:10.3969/j.issn.1674-0904.2019.11.012

**引文格式:** Li TY, Lin ZC, Ge Q, et al. Two indexes for comparison of model prediction effect and their computation in R[J]. J Cancer Control Treat, 2019, 32(11): 1018-1023. [李彤阳, 林卓琛, 葛琪, 等. 两种比较模型预测效果的评价指标及其 R 实现[J]. 肿瘤预防与治疗, 2019, 32(11): 1018-1023.]

## Two Indexes for Comparison of Model Prediction Effect and Their Computation in R

Li Tongyang<sup>#</sup>, Lin Zhuochen<sup>#</sup>, Ge Qi, Chen Wen, Zhang Jinxin

<sup>#</sup>Contributed equally

School of Public Health, Sun Yat-sen University, Guangzhou 510080, Guangdong, China

**Corresponding author:** Zhang Jinxin, E-mail: zhjinx@mail.sysu.edu.cn

This study was supported by National Natural Science Foundation of China (NO. 81773545) and by grants from Bureau of Education in Guangdong Province (NO. 2019-286-12).

**[Abstract]** **Objective:** To introduce net reclassification improvement and integrated discrimination improvement, and clarify the significance and application of these two statistical indexes in the evaluation of prediction effect of logistic regression models commonly used in medical research. **Methods:** Definitions of net reclassification improvement and integrated discrimination improvement were given with examples to illustrate the computation of the two indexes in R. We compared them with commonly used indexes comparing prediction effect such as receiver operating characteristic curve, area under the curve, sensitivity and specificity, so as to elaborate their characteristics in application. **Results:** In logistic regression models, net reclassification improvement and integrated discrimination improvement could quantify the improvement of prediction effect from old to new models. Estimated values and hypothesis testing results of this improvement can be easily obtained in R software. **Conclusion:** Net reclassification improvement and integrated discrimination improvement quantitatively reflect the difference before and after the

model optimization, and have practical value. R software can be used to complete the calculations.

**[Key words]** Prediction model; Evaluation index; Net reclassification improvement; Integrated discrimination improvement

**[收稿日期]** 2019-07-21 **[修回日期]** 2019-10-31

**[基金项目]** \* 国家自然科学基金(编号:81773545); 2019 年度广东省高等教育教学改革项目(编号:2019-286-12)

**#共同第一作者**

**[通讯作者]** <sup>△</sup> 通讯作者: 张晋昕, E-mail: zhjinx@mail.sysu.edu.cn

在医学研究特别是临床医学研究中, Logistic 回

归是常见的分析方法<sup>[1-2]</sup>。以两分类 logistic 回归为例,模型的因变量为患病与否,自变量为一系列与该病相关的指标。通过建立 Logistic 回归模型我们可以通过每个观察对象的疾病相关指标,计算出其患病预测概率。目前,Logistic 回归在临床肿瘤患病风险研究中应用广泛:如基因多态性与食管癌患病风险关系研究<sup>[3]</sup>,2 型糖尿病并发恶性肿瘤风险研究<sup>[4]</sup>,以及卵巢肿瘤良恶性的鉴别诊断<sup>[5]</sup>,也有癌症术后不良反应发生率及其相关危险因素的研究等<sup>[6-7]</sup>。

在实际研究中,如果涉及到两个 Logistic 模型预测效果的比较,例如两种诊断试验的比较,或是在常规影响因素中添加某一新的生物标志物或感兴趣的危险因素,我们常比较两个模型的灵敏度、特异度、准确度、受试者工作特征(receiver operating characteristic,ROC)曲线<sup>[8]</sup>以及曲线下面积(area under curve,AUC)<sup>[9-10]</sup>。通常,Logistic 回归会给出模型的分表,据此可以看到两个模型各自的灵敏度、特异度、准确度等指标,若是想进一步整体比较两模型孰优孰劣,则需比较 AUC。AUC 综合了 ROC 曲线上所有点作为诊断截点的情况,直观地反映试验的整体预测效果,在临床上尤其是诊断试验的比较中应用广泛。但在实际应用中,我们常常更关注某一诊断截点下两个模型对个体的分类能力,这时 AUC 就不那么好用了;此外,AUC 的敏感性有时并不强,其增量并不明显,专业意义也不易理解<sup>[11-12]</sup>。如果我们关心新模型相比于旧模型,对于个体的预测概率改善了多少,或是判断新加入的生物标志物是否影响到临床决策,那么只采用常规的方法就不够了。近年来,其他比较新旧模型预测效果的方法相继提出,其中运用最多的是 Pencina 提出的净重新分类指数(net reclassification improvement,NRI)和整体鉴别指数(integrated discrimination improvement,IDI)<sup>[11,13-15]</sup>。NRI 关注的是在诊断截点处,通过考察使用新模型后个体预测概率的变化情况,或个体被重新分类的情况,得出新模型比旧模型使得不同组别个体更有利于分到正确组的概率,并对两组预测概率的改善求和而 IDI 则是使用两个模型预测概率均值的差来整体评价新模型改善的程度,故 NRI 和 IDI 更贴合应用需求<sup>[11,13-14]</sup>。

## 1 NRI 和 IDI 的意义与计算

### 1.1 常用的模型评价指标

在 Logistic 回归模型中,由自变量计算预测概率,当预测概率大于某截点值时,可认为该观察对象

属于 1 组,否则属于 0 组。而根据课题研究背景,常以 1 代表患病( $y = 1$ ),0 代表不患病( $y = 0$ ),此时我们就可以认为预测概率高于该截点时,该观察对象被预测为患病组(反之被预测为不患病组)。

通过诊断试验四格表可以计算出灵敏度、特异度、约登指数等一系列指标。灵敏度为患病组中个体被正确预测为患者的概率,特异度为不患病组中个体被正确预测为健康者的概率,约登指数为灵敏度与特异度之和减去 1。

另外,根据连续型变量(如预测概率)截点大小的选取,得到的灵敏度和特异度也会相应地改变,这种变化通过连线的形式呈现出来就是 ROC 曲线,计算 ROC 曲线下面积则可以得到 AUC 值,它是对连续型变量无固定截点时预测能力的概括性刻画。

### 1.2 NRI 指数

净重新分类指数 NRI<sup>[11,13-15]</sup>,评价的是在两模型采用最优诊断截点进行预测时,与旧模型相比,使用新模型使得个体的预测结果得到改善的概率,包括事件发生组改善概率与不发生组改善概率两部分(在 R 中分别以 NRI+ 和 NRI- 表示)。R 中对于 NRI 指数的计算有两种不同的方式,一是将预测风险概率视为连续变量,对应于 R 中的 diff 方法;一是将风险概率划分为若干区间,对应于 R 中的 category 方法。

1.2.1 diff 法 NRI 指数 diff 法 NRI 指数以某一概率变化值(概率变化的大小,而非概率截点)作为判断新旧模型预测概率变化的依据,只要个体的预测概率变化超过了某一变化值,例如 0.02,则认为该个体在新模型下预测的概率和旧模型的不同。相应的计算公式如下:

$$NRI = P(\text{up}|\text{event}) - P(\text{down}|\text{event}) + P(\text{down}|\text{nonevent}) - P(\text{up}|\text{nonevent}) \quad (1)$$

为了便于理解这个公式,我们以预测结局为患病、不患病两种情况的研究为例,其中 up|event 和 down|event 分别表示患病组中个体采用新模型后预测概率值上升和下降的情形(概率变化的幅度超过了设定值才认为是上升或下降),down|nonevent 和 up|nonevent 表示不患病组中,个体采用新模型后预测的概率值下降和上升的情形(上升和下降的定义同前),并分别对应于各情形发生的概率。在实际计算中,可以使用以下公式计算 NRI 指数:

$$NRI = \frac{\#up\_event - \#down\_event}{\#event} + \frac{\#down\_nonevent - \#up\_nonevent}{\#nonevent} \quad (2)$$

其中#up\_event 和 #down\_event 分别表示患病组中新模型预测的患病概率上升和下降的人数(上升和下降的定义同前),相应的,#up\_nonevent 和 #down\_nonevent 分别表示不患病组中新模型预测的患病概率上升和下降的人数(上升和下降的定义同前)。**#event** 表示患病组人数,**#nonevent** 表示不患病组人数。

由上述公式可知,NRI 指数越大,则与旧模型相比,在新模型中个体无论是本来属于患病组还是不患病组,均有更大机会被分至实际所属的组别,预测结果的正确率得以提高,使用新模型后患病组计算出来的预测概率( $P(y = 1)$ )更高了,而不患病组计算出来的预测概率( $P(y = 1)$ )更低了。

**1.2.2 Category 法 NRI 指数** Category 法 NRI 指数是将预测风险概率以若干个概率截点划分为感兴趣的概率区间,并基于一个重新分类表,考察个体被新模型预测后重新分类到不同概率区间的情况,它不关注个体概率变化了多少,而是直接分析个体发生某事件的预测概率所属的区间类别得到优化的概率。仍以预测结局为患病和不患病的研究为例,当仅设置一个概率截点时,由于实际上是将概率区间二分类了,其重新分类表相当于一个患病组和不患病组分开统计的配对四格表(见表 1),可按公式(3)计算 NRI 指数:

$$NRI = \frac{\#event_1 - \#event_0}{\#event} + \frac{\#nonevent_0 - \#nonevent_1}{\#nonevent} \quad (3)$$

表 1 二分类 category 法 NRI 指数新旧模型预测结果的重新分类表

Table 1. Reclassification Table for Category-Based NRI with One Cut-off

Old model	New model	
Predicted probability	< cut *	≥ cut
Patients		
< cut	a1	b1 (#event1)
≥ cut	c1 (#event0)	d1
Controls		
< cut	a2	b2 (#nonevent1)
≥ cut	c2 (#nonevent0)	d2

\* Cut: Probability cut-off.

结合公式(3)和表 1, **#event<sub>1</sub>** 和 **#nonevent<sub>0</sub>** 分别表示患病组和不患病组中新模型预测正确而旧模型预测错误的人数(**b1** 和 **c2**), **#event<sub>0</sub>** 和 **#nonevent<sub>1</sub>** 分别表示患病组和不患病组中新模型预测错误而旧模型预测正确的人数(**c1** 和 **b2**)。 **#event** 表示患病组人数,**#nonevent** 表示不患病组人数。可见,由于 **a1**, **d1**, **a2**, **d2** 是两种模型预测的概率区间相同的情况,

在计算 NRI 指数时并未用到。

同时不难发现,此时 category 法的 NRI 指数等于新模型灵敏度与特异度之和减去旧模型灵敏度与特异度之和,在数值上也相当于新模型的约登指数减去旧模型的约登指数。

当我们设置多个概率截点时,模型预测的风险概率就被划分成不同的概率区间(规定新旧模型的预测概率区间采用相同的划分方式),此时我们认为患病组和不患病组个体患病风险概率的分布不同,理论上患病组的人更多地集中在风险概率高的区间,不患病组的人更多地集中在风险概率低的区间,因为模型计算出来的预测概率越大,越有利于该个体预测为患病,同理,模型预测的概率越小,越有利于个体被预测为不患病。此时 NRI 指数计算公式的形式与公式(2)相同,含义略有差别:

公式中**#up\_event** 和 **#down\_event** 分别表示患者采用新模型预测后所属的概率区间比旧模型预测的概率区间等级升高和降低的人数, **#up\_nonevent** 和 **#down\_nonevent** 分别表示不患病者采用新模型预测后所属的概率区间比旧模型预测的概率区间等级升高和降低的人数,相应的重新分类表(以设定 2 个概率截点为例)见表 2,其中 **x1**, **x5**, **x9**; **y1**, **y5**, **y9** 表示预测结果相同的情形。实际上,公式(2)用在此处,仅仅是升高和降低的含义发生了改变,在 diff 法中升高和降低的前提是个体预测概率的差值超过了设定的概率变化值,在 category 法中升高和降低的判断是依据个体所属概率区间等级的改变。另外,category 法中只设定一个概率截点的情形也适用于公式(2),相应含义与设定多个概率截点时相同。

表 2 三分类 category 法 NRI 指数新旧模型预测结果的重新分类表

Table 2. Reclassification Table for Category-Based NRI with Two Cut-offs

Old model	New model		
Predicted probability	< cut1	[ cut1, cut2)	≥ cut2
Patients			
< cut1	x1	x2	x3
[ cut1, cut2)	x4	x5	x6
≥ cut2	x7	x8	x9
Controls			
< cut1	y1	y2	y3
[ cut1, cut2)	y4	y5	y6
≥ cut2	y7	y8	y9

\* Cut: Probability cut-off.

**1.2.3 NRI 指数的假设检验** 对于 Category 法 NRI 指数,可以使用公式(4)计算其对应的 **z** 值,实施总体 NRI 指数是否等于 0 的假设检验<sup>[13]</sup>,其中#

up\_event、#down\_event、#up\_nonevent 和 #down\_non-

$$Z_{NRI} = \frac{NRI}{\sqrt{\frac{\#up\_event + \#down\_event}{\#event^2} + \frac{\#down\_nonevent + \#up\_nonevent}{\#nonevent^2}}} \quad (4)$$

而对于 diff 法的 NRI 指数,可使用 Bootstrap(自助)法算出其对应的置信区间和 P 值。

### 1.3 IDI 指数

1.3.1 IDI 指数的定义与计算 整体鉴别指数 IDI<sup>[11,13-14]</sup>,或称综合判别改善指数,它考虑了不同诊断截点的情况,并且从平均概率的角度反映了模型的整体改善状况。计算公式如下:

$$IDI = (\overline{P_{event,new}} - \overline{P_{event,old}}) - (\overline{P_{nonevent,new}} - \overline{P_{nonevent,old}}) \quad (5)$$

仍以预测结局为患病、不患病的研究为例,其中  $\overline{P_{event,new}}$  是新模型对患病组所有个体预测患病概率的均值,  $\overline{P_{event,old}}$  是旧模型对患病组所有个体的预测患病概率的均值,  $\overline{P_{nonevent,new}}$  是新模型对不患病组所有个体预测患病概率的均值,  $\overline{P_{nonevent,old}}$  是旧模型对不患病组所有个体预测患病概率的均值。

由公式(5)可以发现,基于预测概率越高越可能为患病组的设定, IDI 指数越大,说明新模型可使两组个体被判定为本应所属类别的指向性更明确,鉴别能力得以提高。

1.3.2 IDI 指数的假设检验 IDI 指数可以使用公式(6)计算其对应的 z 值,实施总体 IDI 指数是否等于 0 的假设检验<sup>[13]</sup>,其中, SE#event 是患病组中  $\overline{P_{event,new}} - \overline{P_{event,old}}$  的标准误, SE#nonevent 指的是不患病组中  $\overline{P_{nonevent,new}} - \overline{P_{nonevent,old}}$  的标准误。

$$Z_{IDI} = \frac{IDI}{\sqrt{(SE\#event)^2 + (SE\#nonevent)^2}} \quad (6)$$

## 2 实例分析

本文采用《SPSS 8.0 for Windows 统计软件使用指南》<sup>[16]</sup>中 Logistic 回归分析章节 50 例急性淋巴细胞白血病病人资料(文中例 14.3)。数据共包含 5 个变量,其含义分别为: X1:入院治疗时取外周血中白细胞数(千个/mm<sup>3</sup>); X2:浸润淋巴结分级(根据浸润程度分 0,1,2 三级); X3:治疗缓解后病人有无巩固治疗(有=1,无=0); t:随访病人生存时间,按月计算;研究结局 Y:病人生存时间是否小于 1 年(一年或以上=0,一年以下=1)。本次研究目的是建立合适的 Logistic 回归模型,对淋巴细胞白血病一年内是否死亡进行预测。现分别建立两个模型,其

event 取 category 法使用公式(2)时对应的含义。

中模型 1 纳入自变量入院治疗时取外周血中白细胞数和浸润淋巴结分级,模型 2 在模型 1 自变量的基础上再加上治疗缓解后病人有无巩固治疗,模型建立后使用 NRI 与 IDI 对两模型预测效果进行评价。

在 R 中可使用 nricens 包和 PredictABEL 包进行指标的计算,相应的 R 代码如下:

```
#读入需要使用的包
#install.packages("PredictABEL")
# install.packages("nricens")
library(PredictABEL)
library(nricens)
#建立计算最佳截点函数
opt.cut = function(perf, pred) {
  cut.ind = mapply(FUN = function(x, y, p) {
    d = (x - 0)^2 + (y - 1)^2
    ind = which(d == min(d))
    c(sensitivity = y[[ind]], specificity = 1 - x[[ind]]),
    cutoff = p[[ind]])
  }, perf@x.values, perf@y.values, pred@cutoffs)
}
#读取数据
setwd("D:\\data")
dat <- read.csv("data.csv", na.strings = "#NULL!")
#拟合模型
mod1 <- glm(y ~ X1 + X2, family = binomial(link = 'logit'), data = dat)
mod2 <- glm(y ~ X1 + X2 + X3, family = binomial(link = 'logit'), data = dat)
pre1 <- predict(mod1, newdata = dat, type = 'response')
pre2 <- predict(mod2, newdata = dat, type = 'response')
pred1 <- prediction(pre1, dat$y); perf1 <- performance(pred1, "tpr", "fpr")
pred2 <- prediction(pre2, dat$y); perf2 <- performance(pred2, "tpr", "fpr")
cutpoint1 <- opt.cut(perf1, pred1)[3]; cutpoint2 <- opt.cut(perf2, pred2)[3]
pre_case1 <- pre1 > cutpoint1; pre_case2 = pre2 > cutpoint1
table(dat$y, pre_case1); table(dat$y, pre_case2)
#使用 nricens 包计算 NRI 指数
NRIa <- nrabin(event = dat$y, p.std = pre1, p.new = pre2, updown = "category", cut = cutpoint2, niter = 10000, alpha = 0.05)
NRIb <- nrabin(event = dat$y, p.std = pre1, p.new = pre2,
```

```
updown = "diff", cut = 0, miter = 10000,
alpha = 0.05)
#使用 PredictABEL 包计算 NRI 指数和 IDI 指数
reclassification( data = dat, cOutcome = 5, predrisk1 = pre1,
predrisk2 = pre2, cutoff = c(0, cut-
point2, 1))
```

结果如表 3 和表 4:

表 3 基于是否考虑“有无巩固治疗”所得两模型预测概率的重新分类表

Table 3. Predictive Probabilities of Two Models With or Without the Consolidation Therapy Variable

Model 1	Model 2	
Predictive probability	<0.825	≥0.825
Death within 1 year (n = 30)		
<0.825	8	21
≥0.825	0	1
Survival >1 year (n = 20)		
<0.825	18	2
≥0.825	0	0

本例中选取的概率截点 0.825, 是通过拟合模型预测概率的 ROC 曲线, 并取约登指数最大值对应的概率确定的。如果已经确定好概率截点, 可直接给出相应的赋值。表 3 相当于二分类 category 法 NRI 指数的重新分类表。读者也可据表 3 按公式 (3) 计算 NRI (category), 结果与表 4 中软件计算的表 4 基于是否考虑“有无巩固治疗”所得两个统计模型比较

Table 4. Two Statistical Models With or Without the Consolidation Therapy Variable

Software package	NRI (category)	P	NRI (diff)	P	IDI	P
nricens	0.600	$2.00 \times 10^{-8}$	1.167	$1.95 \times 10^{-07}$	— *	— *
PredictABEL	0.600	$1.11 \times 10^{-8}$	1.167	$1.39 \times 10^{-07}$	0.298	$2.93 \times 10^{-6}$

NRI: Net reclassification improvement; IDI: Integrated discrimination improvement.

\* No corresponding results provided.

### 3 讨论

生物医学数据中经常用到 Logistic 回归模型来对疾病发生进行预测。除常见的灵敏度、特异度、约登指数、AUC 等指标外, NRI 指数和 IDI 指数也可用于评价模型预测的效果, 且是全面的定量分析指标。R 软件提供的 nricens 包和 PredictABEL 包能方便地实现计算。

ROC 曲线与 AUC 值刻画了模型在选用不同诊断截点时的灵敏度和特异度, AUC 的提高通常可作为模型预测能力改善的第一判断标准, 然而实际应用中, 当两个模型的 AUC 曲线存在交叉时, 或 AUC 的增量极小时, 仅仅采用 AUC 可能就很难说明问题了, 此时 NRI 和 IDI 指数则是更好的选择。当设置了一至多个概率截点, 关心截点处的分类情况时, category 法的 NRI 指数就更合适, 如果目前还无法

值相同。此外, 结果报告时还应给出一年内死亡与未死亡组各自的概率的改善<sup>[15]</sup>, 这与 R 输出组与 NRI + 和 NRI - 对应(此处以略)。

由表 4 可知, 两个 R 软件包计算的 NRI 指数结果一致, nricens 包没有提供 IDI 的计算结果; 实际上两个软件包都会给出相应的置信区间, 此处受篇幅限制未列出。由表 4 可知, 模型 2 (包含“有无巩固治疗”的模型) 与模型 1 (不包含“有无巩固治疗”的模型) 相比, 得到的 NRI 指数和 IDI 指数均为正数, 即模型 2 的分类能力优于模型 1, 也即, 包含“有无巩固治疗”的模型比不包含的模型分类能力好。具体好在何处? 以 category 法 NRI 指数为例, 值为 0.600, 说明使用模型 2 相对于模型 1, 使得一年内死亡和未死亡的患者更有利于得到正确分类的概率之和为 0.600 (注意, 此处不可理解或表示为百分比或概率<sup>[15]</sup>),  $P < 0.001$ , 结果具有统计学意义; 类似地, IDI 指数是 0.298, 说明采用模型 2 相比于模型 1, 在一年内死亡组 ( $Y = 1$ ) 中预测患者一年内死亡的概率均值的提高量和在一年内未死亡组 ( $Y = 0$ ) 预测患者一年内死亡的概率均值的降低量之和为 0.298, 说明模型 2 预测能力比模型 1 要好,  $P < 0.001$ , 可见结果具有统计学意义。

确定明确的截点, 或需整体评价两模型的预测能力, 则 IDI 指数和 AUC 更适合<sup>[11]</sup>。

本文中虽然给出了两种计算 NRI 指数的方法, 但是实际应用中通常采用 category 法, 即给出概率截点划分预测概率区间, 并以此作为个体分类的依据。由于 category 法 NRI 指数的计算与设定的概率截点有关, 因此, 预测概率区间的划分应该选择适宜的界值, 需要结合临床专业的需求来确定。区间划分不恰当时, 可能得不到统计上有意义的结论, 或者所得结论缺少临床实践的指导意义<sup>[11]</sup>; 同时, 由于计算时只考虑了已调变化, 未考虑变化的等级数, 因此可能存在信息缺失<sup>[12]</sup>; 这可能是 NRI 指数使用存在的缺陷。此外, 在模型比较时, 需注意概率截点选择的侧重点是否一致, 比如有的模型以提升特异度为侧重点, 有的模型以提升灵敏度为侧重点, 这就需要在 NRI 和 IDI 指数的过程时考虑两个模型的

“侧重类型”是否相同,侧重类型不同的模型间没有可比性。当前研究大多数研究考虑灵敏度和特异度各占一半的比重,但某些特殊的模型例如用于初筛的模型则偏重灵敏度,因此在模型的评价上需更加谨慎的应用 NRI 和 IDI 指数<sup>[17]</sup>。对于 IDI 指数来说,在计算总改善情况时,通常患病组和不患病组占有同等的比重,但对于侧重于灵敏度的模型,计算时应给予患病组概率改变量更多的权重。对此类模型,可以使用调整过权重后的 IDI。有兴趣的读者可做进一步的研究<sup>[18]</sup>。

理论上,除了将 NRI 和 IDI 指数应用在 Logistic 回归模型外,也可应用于其他可以输出预测概率的模型,如决策树、随机森林、神经网络等。但是对于这两个指标在这些模型中的适用性,仍需进一步的研究与讨论。

综上,NRI 和 IDI 指数是近年来新兴的模型比较评价指标,在模型比较中能够轻松处理 AUC 等不能解决的复杂情形,具有独特的优势,但二者在肿瘤流行病学研究中应用还较少。本文结合临床肿瘤风险预测实例,详细阐述了两种指标的计算原理和适用情况,以及使用 R 语言计算两种指标估计值及假设检验的方法,以期为临床肿瘤研究工作者提供更为详尽的参考。当涉及两模型优劣的比较,或考察某一因素的加入对模型预测效果的改善程度时,除了使用约登指数及 AUC 外,建议研究者不妨同时考虑 NRI 和 IDI 指数,以期更加全面地展示模型的改善情况。

**作者声明:**本文全部作者对于研究和撰写的论文出现的不端行为承担相应责任;并承诺论文中涉及的原始图片、数据资料等已按照有关规定保存,可接受核查。

**学术不端:**本文在初审、返修及出版前均通过中国知网(CNKI)科技期刊学术不端文献检测系统的学术不端检测。

**同行评议:**经同行专家双盲外审,达到刊发要求。

**利益冲突:**所有作者均声明不存在利益冲突。

**文章版权:**本文出版前已与全体作者签署了论文授权书等协议。

#### [参考文献]

[1] Molinari M, Ayloo S, Tsung A, et al. Prediction of perioperative mortality of cadaveric liver transplant recipients during their evalu-

ations [J]. *Transplantation*, 2019, 103(10):e297-e307.

- [2] Nagata T, Ohara T, Hata J, et al. NT-proBNP and risk of dementia in a general Japanese elderly population: The Hisayama study [J]. *J Am Heart Assoc*, 2019, 8(17):e011652.
- [3] 刘芳,谭榜亮,彭小东. 四川北部地区 AGT, XRCC1 基因多态性与食管癌风险的关系[J]. *肿瘤预防与治疗*, 2016, 29(1):5-11.
- [4] 秦娟,吴艳,朱晓荣,等. 2 型糖尿病并发恶性肿瘤风险的 Logistic 回归分析 [J]. *肿瘤预防与治疗*, 2017, 30(6):433-436,452.
- [5] 陈静,曾令玲,罗光辉,等. 超声征象评价在卵巢恶性肿瘤鉴别诊断价值的 Logistic 回归分析 [J]. *肿瘤预防与治疗*, 2019, 32(2):139-143.
- [6] 罗霜,蔡博文,刘进,等. 经鼻蝶垂体瘤手术后恶心呕吐的发生率及危险因素的 Logistic 回归分析 [J]. *肿瘤预防与治疗*, 2018, 31(4):268-272.
- [7] 樊炜,李进,胡智洪,等. 经鼻蝶垂体瘤术后患者腺垂体功能低下发生情况及相关危险因素分析 [J]. *肿瘤预防与治疗*, 2018, 31(3):173-177.
- [8] Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction [J]. *Circulation*, 2007, 115(7):928-935.
- [9] 宋花玲,贺佳,虞慧婷,等. 应用 ROC 曲线下面积对两相关诊断试验进行评价和比较 [J]. *第二军医大学学报*, 2006, 27(5):562-563.
- [10] 汪云云,陈松灿. 基于 AUC 的分类器评价和设计综述 [J]. *模式识别与人工智能*, 2011, 24(1):64-71.
- [11] 于莉莉,武颂文,夏结来. 利用净重新分类指数与整体鉴别指数评价一种新危险因素的补充预测能力 [J]. *中国卫生统计*, 2017, 34(5):761-763.
- [12] 周立志. 一种新的风险等级预测的模型改善评价方法 [D]. 广州:南方医科大学, 2015.
- [13] Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, et al. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond [J]. *Stat Med*, 2008, 27(2):157-172.
- [14] Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers [J]. *Stat Med*, 2011, 30(1):11-21.
- [15] Leening MJ, Vedder MM, Witteman JC, et al. Net reclassification improvement: Computation, interpretation, and controversies: a Literature review and clinician's guide [J]. *Ann Intern Med*, 2014, 160(2):122-131.
- [16] 刘润幸. SPSS 8.0 for Windows 统计软件使用指南(第一版) [M]. 湛江:广东人民出版社, 1999:209-214.
- [17] Chi YY, Zhou XH. The need for reorientation toward cost-effective prediction: Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by Pencina et al., *Statistics in Medicine* [J]. *Stat Med*, 2008, 27(2):182-184.
- [18] Kraemer HC. Comments on 'Evaluating the added predictive ability of a new marker' by M. Pencina, R. D'Agostino, R. D'Agostino Jr, R. Vasan, *Statistics in Medicine* [J]. *Stat Med*, 2008, 27(2):196-198.