·计算机应用·

净重新分类指数和综合判别改善指数的应用及SAS实现*

朱博文[1] 王 威[2] 王星云[1] 郭锡熔[1△]

【提 要】 **目的** 利用SAS宏程序实现净重新分类指数(net reclassification improvement,NRI)、综合判别改善指数(integrated differentiation improvement,IDI)及其95%可信区间的计算,探讨两指标在疾病预测模型效能评价中的应用特征。**方法** 以SAS软件内置冠心病预测数据集为实例,分别构建新旧logistic回归预测模型,提取个体发病预测概率,区分事件组与非事件组,采用Z检验计算NRI、IDI及95%可信区间,同时探讨小样本条件下T检验的适用场景。**结果** 本研究共计纳入5209例观察对象,其中冠心病事件组1449例,非事件组3760例。通过SAS宏程序运算结果显示,在原有模型基础上引入超重变量后,NRI为0.240,95% *CI*:0.187~0.293,*P*<0.001;IDI为0.003,95% *CI*:0.002~0.005,*P*<0.001,提示新增变量后的新模型预测效能优于旧模型。**结论** 基于SAS宏可实现NRI与IDI的统计运算。实际分析中应重视样本量条件,结合综合评判模型增益,可用于临床预测模型的效能比较。

【关键词】 净重新分类指数 综合判别改善指数 SAS宏
【中图分类号】 R195.1 【文献标志码】 A **DOI** 10.11783/j.issn.1002-3674.2026.05.029

净重新分类改善(net reclassification improvement,NRI)指数和综合判别改善(integrated differentiation improvement,IDI)指数是2008年由Pencina等提出的比较两模型预测能力优劣的统计学指标[1]。常见临床研究中,想要在原有模型中引入新指标观察模型的预测能力能否提高时,仅凭受试者工作特征曲线(receiver operating characteristic,ROC)下面积的增减很难评价其是否有意义。此时,NRI和IDI可以弥补临床研究中ROC曲线的短板:NRI可以在某个诊断截点处,通过计算新、旧模型个体预测概率的变化及重新分类情况,得出新、旧模型的个体在不同组别间被分到正确组的概率[2]。不同的是,NRI考虑的是某截断值下的新模型改善情况,IDI从整体预测概率角度计算模型的优劣,即在不牺牲平均特异性的情况下衡量新模型对平均灵敏度的改进[3]。近年来,NRI和IDI作为一种综合反映模型预测能力的指标,越来越受研究者们关注,多数研究中两指标实现的统计软件主要是R语言[4-5]。因此,本文通过实例介绍SAS宏程序计算NRI和IDI的步骤及结果解释,并展示两指标的应用特点。

实例分析

1. 净重新分类改善指数

NRI的计算需要构建一个重分类表,首先根据新、旧模型计算事件的发生概率,得到的风险概率按等级划分,并将新、旧模型概率的变化用交叉表形式列出,经典NRI可信区间算法根据事件组和非事件组中两模型预测概率的差异是否服从正态分布Z检验的统计思想,计算方法为:

$$NRI=(P_{提高,事件组}-P_{下降,事件组})-(P_{提高,非事件组}-P_{下降,非事件组})$$

$P_{提高,事件组}$:事件组提高事件数/事件组总例数;$P_{降低,事件组}$:事件组降低事件数/事件组总例数;$P_{提高,非事件组}$:非事件组提高事件数/非事件组总例数;$P_{降低,非事件组}$:非事件组降低事件数/非事件组总例数。

假定事件组与非事件组个体相互独立,得到一个关于无效假设$NRI=0$的检验方法,样本量较大时,统计量Z值服从标准正态分布。

$$Z=\frac{NRI}{\sqrt{\left(\frac{P_{提高,事件组}+P_{降低,事件组}}{事件组例数}\right)+\left(\frac{P_{提高,非事件组}+P_{降低,非事件组}}{非事件组例数}\right)}}$$

同时计算NRI的可信区间:

NRI_STDERR=SQRT(((($P_{提高,事件组}+P_{降低,事件组}$)/事件组例数^2-($P_{提高,事件组}-P_{降低,事件组}$)^2/事件组例数^3+(($P_{提高,非事件组}+P_{降低,非事件组}$)/非事件组例数^2-($P_{降低,非事件组}+P_{提高,非事件组}$)^2/非事件组例数^3))。

LOW_NRICI=NRI-1.96×NRI_STDERR;

UP_NRICI=NRI+1.96×NRI_STDERR。

2. 综合判别改善指数

经典IDI及可信区间算法根据事件组和非事件组中两个模型预测概率的差异是否服从正态分布Z检验的统计思想。值得注意的是,当模型比较的样本量较小时,经典的IDI检验不能有效处理小样本量模型比较的情况,可采用T检验。IDI是基于模型对个体的预测概率计算所得,计算方法为:

$$IDI=(P_{新模型,事件组}-P_{旧模型,事件组})-(P_{新模型,非事件组}-P_{旧模型,非事件组})$$

$P_{新模型,事件组}$:事件组中新模型对每个人预测概率的

*基金项目:国家重点研发计划资助项目(2021YFC2701900;2021YFC2701903)
1.上海交通大学医学院附属同仁医院虹桥国际研究院(200336)
2.南京中医药大学附属医院GCP中心
△通信作者:郭锡熔,E-mail:xrguo@shsmu.edu.cn

平均值；$P_{旧模型,事件组}$：事件组中旧模型对每个人预测概率的平均值；$P_{新模型,非事件组}$：非事件组中新模型对每个人预测概率的平均值；$P_{旧模型,非事件组}$：非事件组中旧模型对每个人预测概率的平均值。

事件组中预测概率越高模型越准确，当$P_{新模型,事件组}$与$P_{旧模型,事件组}$相减时，值越大表明新模型预测能力越好。而非事件组中预测概率越低模型越准确，当$P_{新模型,非事件组}$与$P_{旧模型,非事件组}$相减时，值越小表明新模型预测能力越好。因此两组差异再相减，值越大表明新模型预测能力越高。如果IDI为负值，说明旧模型优于新模型；如果IDI为正值，说明新模型优于旧模型。计算统计量Z值是否服从标准正态分布，以判断IDI与0相比是否具有显著性差异。

$$Z = \frac{IDI}{\sqrt{(SE_{事件组})^2 + (SE_{非事件组})^2}}$$

其中，$SE_{事件组}$代表事件组新、旧模型预测概率平均值差异的标准误，而$SE_{非事件组}$代表非事件组新、旧模型预测概率平均值差异的标准误。

同时计算IDI的可信区间：

$$LOW_IDICI = IDI - 1.96 \times \sqrt{(SE_{事件组})^2 + (SE_{非事件组})^2}$$

$$UP_IDICI = IDI + 1.96 \times \sqrt{(SE_{事件组})^2 + (SE_{非事件组})^2}$$

基于小样本量人群新、旧模型比较时，需要考虑事件组和非事件组方差齐性，则采用T检验。

结　果

1. 数据来源

为介绍SAS宏程序的应用，本研究用一个SAS Help自带的测试数据（名为“Heart”）展示NRI和IDI及其可信区间的计算过程。该数据集提取于弗明翰心血管疾病风险研究的数据，主要包括年龄、收缩压、舒张压、血脂、身高、体重、吸烟和冠心病情况等变量，旨在对人群冠心病预测能力进行比较。此测试数据集含5209个观察值，每个观测值有17个变量，图1展现了该数据集的前10个观测值。其中，高密度脂蛋白、血尿酸和年龄为连续型变量，慢性肾脏病、心血管疾病史和高血压为分类变量。我们想要分析与年龄、性别、吸烟状况和高脂血症的冠心病风险预测模型相比，新模型变量添加超重后预测能力能否改善。

编号	冠心病诊断	性别	基线年龄	吸烟	饮酒	血压	超重
1	0	0	29	0	0	0	1
2	0	0	41	0	0	0	1
3	0	0	57	1	1	1	1
4	0	0	39	0	1	0	1
5	0	1	42	1	1	0	1
6	0	0	58	0	0	1	1
7	0	0	36	1	0	0	1
8	0	1	53	0	1	0	0
9	0	1	35	0	0	0	1
10	0	1	52	1	1	0	0

图1　输入数据集（Want）的前10个观察值

2. SAS实现与结果

（1）NRI指数

①定义年龄、性别、吸烟状况和高脂的冠心病风险预测模型为旧模型，添加超重因素后风险预测模型为新模型。计算冠心病组新、旧模型的预测阳性概率和非冠心病组新、旧模型的预测阳性概率。SAS程序如下：

```
proc logistic desc data=want;
model CHD_diag = Sex AgeAtStart Smoking Chol_Status BP_Status;
output out=test1 p=prob1;
run;
proc logistic desc data=want;
model CHD_diag = Sex AgeAtStart Smoking Chol_Status BP_Status Weight_Status;
output out=test2 p=prob2;
run;
data want;
merge test1 test2;
run;
```

②提取事件组和非事件组中新、旧模型提高或降低的事件数，SAS 程序如下：

```
data nri_inf;
set want;
pdiff=prob2-prob1;
if CHD_diag=1 then do;
down_event=(pdiff<0); up_event=(pdiff>0);
down_nonevent=0;up_nonevent=0;
end;
if CHD_diag=0 then do;
down_nonevent=(pdiff<0); up_nonevent=(pdiff>0);down_event=0;up_event=0;
end;
proc sql;
create table nri_inf1 as
select sum(up_nonevent), sum(down_nonevent), sum(up_event),sum(down_event)
into : num_nonevent_up_user, : num_nonevent_down_user, :num_event_up_user, :num_event_down_user
from nri_inf;
quit;
```

结果输出如图 2 所示：

Obs	_TEMG001	_TEMG002	_TEMG003	_TEMG004
1	2437	1323	1113	336

图 2 冠心病组和非冠心病组提高或下降的病例数

③运用 SAS 9.4 进行 SAS 宏的开发，将前期计算的事件组提高或下降病例数，非事件组提高或下降的病例数，事件组例数和非事件组例数赋值给宏变量，进而计算新、旧模型预测概率的变化率，根据 Z 检验计算 NRI 值的可信区间。

```
%let n_t=3760;
%let n_r=1449;
%let up_event=_TEMG003;
%let down_event=_TEMG004;
%let up_nonevent=_TEMG001;
%let down_nonevent=_TEMG002;
data nri1; set nri_inf1;
n1=&n_t. ;n2=&n_r. ;
up_event=&up_event. ;
down_event=&down_event. ;
down_nonevent=&down_nonevent. ;
up_nonevent=&up_nonevent. ;
p_up_event=up_event/n2;
p_down_event=down_event/n2;
p_up_nonevent=up_nonevent/n1;
p_down_nonevent=down_nonevent/n1;
nri= (p_up_event-p_down_event) - (p_up_nonevent-p_down_nonevent);
nri_stderr=sqrt((((up_event+down_event)/n2**2-(up_event-down_event)**2/n2**3)+
((up_nonevent+down_nonevent)/n1**2-(down_nonevent-up_nonevent)**2/n1**3));
low_nrici=round(nri-1.96*nri_stderr,.0001);
up_nrici=round(nri+1.96*nri_stderr,.0001);
z_nri=nri/sqrt ( ( (p_up_event+p_down_event)/1449)
+((p_up_nonevent+p_down_nonevent)/3760));
pvalue_nri=2*(1-PROBNORM(abs(z_nri)));
event_correct_reclass=p_up_event-p_down_event;
nonevent_correct_reclass=p_down_nonevent-p_up_nonevent;
z_event=event_correct_reclass/sqrt((p_up_event+p_down_event)/n2);
pvalue_event=2*(1-probnorm(abs(z_event)));
z_nonevent=nonevent_correct_reclass/sqrt((p_up_nonevent+p_down_nonevent)/n1);
pvalue_nonevent=2* (1-probnorm (abs (z_nonevent)));
run;
```

结果输出如图 3 所示：

Obs	_TEMG001	_TEMG002	_TEMG003	_TEMG004	n1	n2	up_event	down_event	down_nonevent
1	2437	1323	1113	336	3760	1449	1113	336	1323

up_nonevent	p_up_event	p_down_event	p_up_nonevent	p_down_nonevent	nri	nri_stderr	low_nrici
2437	0.76812	0.23188	0.64814	0.35186	0.23996	0.027098	0.1868

up_nrici	z_nri	pvalue_nri	event_correct_reclass	nonevent_correct_reclass	z_event	pvalue_event	z_nonevent
0.2931	7.76035	8.44E-15	0.53623	-0.29628	20.4121	0	-18.1673

pvalue_nonevent
0

图 3　NRI值及其可信区间

（2）IDI指数

①计算疾病组新、旧模型的预测阳性概率和非疾病组新、旧模型的预测阳性概率，SAS程序同前。

② 计算疾病组和非疾病组中新、旧模型预测阳性概率的差值和标准差。SAS程序如下：

```
data want1;
set want;
difference1=prob2-prob1;
run;
data want2; set want1; keep difference1 prob2 prob1 CHD_diag;run;
proc sort data=want2; by CHD_diag;run;
proc means mean data=want2 std stderr t probt;
var difference1;by CHD_diag;
output out=result2;
run;
```

结果输出如图4和5所示：

分析变量 difference1				
均值	标准差	标准误差	t 值	Pr>|t|
-0.000857069	0.0250626	0.000408726	-2.10	0.0361

图 4　非冠心病组差异均值、标准差、标准误差、T值和P值

分析变量 difference1				
均值	标准差	标准误差	t 值	Pr>|t|
0.0022265	0.0263064	0.000691077	3.22	0.0013

图 5　冠心病组差异均值、标准差、标准误差、T值和P值

③运用 SAS 9.4 进行SAS宏的开发，将前期计算的两组均数差，标准差和样本例数赋值给宏变量，根据双独立样本均值Z检验计算IDI值及可信区间。

```
%let n_t=3760;
%let n_r=1449;
%let mean_t=-0. 000857069;
%let mean_r=0. 0022265;
%let s_t=0. 0250626;
%let s_r=0. 0263064;
data example1;
```

```
n1=&n_t. ;n2=&n_r. ;
mean1=&mean_t. ;mean2=&mean_r. ;
s1=&s_t. ;s2=&s_r. ;
Z= (mean2-mean1)/sqrt ( (s1**2/n1) + (s2**2/n2));
Pli=(mean2-mean1) -1. 96*sqrt ( (s1**2/n1) + (s2**2/n2));
Plm=(mean2-mean1) +1. 96*sqrt ( (s1**2/n1) + (s2**2/n2));
v=n1+n2-2;
pvalue=2*(1-probnorm(abs(Z)));
diff=mean2-mean1;
run;
```

结果输出如图6所示：

Obs	*n*1	*n*2	mean1	mean2	*s*1	*s*2	*Z*
1	3760	1449	-0.000857069	0.0022265	0.025063	0.026306	3.84055

Pli	Plm	v	*P*value	diff
0.001509888	0.00465725	5207	0.000122761	0.003083569

图 6 IDI值及可信区间计算结果

④基于小样本量人群新、旧模型比较时，采用SAS Proc ttest语句实现：

```
proc ttest data=want1;
var difference1;class CHD_diag; run;
```

考虑到研究事件组中预测事件发病率低，我们基于“want”数据集，按照CHD组和非CHD组1∶1随机抽取30例观察值作为示例。事件组新、旧模型的预测阳性概率和非事件组新、旧模型的预测阳性概率的计算前面已经展示，结果输出如图7所示：

CKD_New	方法	均值	95% 置信限均值		标准差	95% 置信限标准差	
0		-0.00086	-0.00166	-0.00006	0.0251	0.0245	0.0256
1		0.00223	0.000871	0.00358	0.0263	0.0254	0.0273
差 (1-2)	汇总	-0.00308	-0.00462	-0.00154	0.0254	0.0249	0.0259
差 (1-2)	Satterthwaite	-0.00308	-0.00466	-0.00151			

图 7 小样本量人群新、旧模型的预测阳性概率计算

讨 论

净重新分类指数与综合判别改善指数一般用于估算添加新指标后，模型预测效果的改善情况，为疾病预测模型的优劣比较提供证据。预测模型指标的选择需要综合考虑临床可行性，传统的受试者工作特征曲线下方面积是一个基于排名的统计量，用于计算某截断值下预测概率更高[6]。如果模型评估重点在于给定截断值后预测事件的提高或下降水平，那么净重新分类指数可以优选。如果模型的评估重点在于模型的整体改变情况，那么计算不同切点下整体预测概率平均距离则优选综合判别改善指数[7]，因此两指标的选择应谨慎应用。

近年来，重分类方法受到越来越多临床研究者的关注。局部晚期胃癌、造影剂相关肾损伤，肥胖性高血压、死亡预后的研究中均有利用此统计方法的相关文献[8-10]。净重新分类指数与综合判别改善指数应用广泛，一项关于心血管疾病的研究中，在男、女性人群中基于弗明翰风险评分添加心肺适能指标，并应用两指数对模型进行评估后发现能更好地预测心血管疾病风险[11]。也有一项基于住院人群的研究创建内部和外部测试集开发基于CT图像的深度学习辅助诊断模型借助综合判别改善指数进行诊断模型的综合评估[12-13]。前期多数文献中综合判别改善指数的结果解释类似，即事件组和非事件组中两模型预测概率的总差异，但对净重新分类指数的解释普遍不统一。一些研究将总体净重新分类指数解释为正确重新分类的百分比[14-16]。也有研究提出，总净重新分类指数由事件和非事件两个分母组成的分数之和，因此净重新分类指数不应该以一个百分比来表示，而应该作为一个

无单位的统计数据[17]。提供重新分类表将更好地解释汇总的重新分类统计数据。

实际工作中常用R软件和SAS软件实现重分类指数计算，两个软件的编程思路主要是：① 构建模型；② 计算平均预测概率；③ 计算分类结果；④ 检验分类差异。该编程思路存在一定局限：判断净重新分类指数与综合判别改善指数是否有意义基于统计量Z值是否服从正态分布，以判断综合判别改善指数与0相比是否显著性差异。如果预测模型样本量较小时，则需考虑事件组和非事件组预测概率的方差齐性等问题。除了样本量的问题，使用综合判别改善指数时，也需要考虑除模型纳入变量以外的混杂因素。临床实际需求中，当需要评估的预测模型变量数有限，如果研究者认为可能存在混杂因素但考虑卫生经济等因素不适用于包含在模型中，模型评估时可以调整混杂因素，从而使得模型预测解释更加准确。其次，如果预测模型的疾病结局本身发病率低，那么事件组和非事件组预测概率随之较低，最终计算所得综合判别改善指数即使存在统计学差异，但对模型改善的程度仍无法衡量，此时可能需要两个指标进行补充评估。

综上所述，利用SAS宏程序编写计算重分类指标具有操作简单、计算快捷、结果清晰、解释有力等特点，可以作为有效的工具来评估疾病/健康状态与暴露因素间的危险效应。当前重分类指标的评估多用于评估临床研究中增加某一指标后，预测模型的改善程度。未来研究可考虑将NRI和IDI的评估作用与影像学、AI智能诊断和成本效益分析的卫生决策研究相结合。

参 考 文 献

[1] 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.

[2] 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.

[3] Alba AC, Agoritsas T, Walsh M, et al. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature [J]. JAMA, 2017, 318(14):1377-1384.

[4] He XY, Gao YF, Huang YP, et al. Development of a prognostic prediction model for patients with cutaneous malignant melanoma: a study based on the SEER database [J]. Eur Rev Med Pharmacol Sci, 2022, 26(24):9437-9446.

[5] Ran B, Gong J, Shang J, et al. Development and validation of nomograms for predicting survival in differentiated thyroid cancer patients with or without radioiodine therapy [J]. Front Oncol, 2023, 13:1054594.

[6] Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures [J]. Epidemiology, 2010, 21(1):128-138.

[7] Cook NR, Demler OV, Paynter NP. Clinical risk reclassification at 10 years [J]. Stat Med, 2017, 36(28):4498-4502.

[8] Cui Y, Zhang J, Li Z, et al. A CT-based deep learning radiomics nomogram for predicting the response to neoadjuvant chemotherapy in patients with locally advanced gastric cancer: A multicenter cohort study [J]. EClinicalMedicine, 2022, 46:101348.

[9] Roscigno G, Quintavalle C, Biondi-Zoccai G, et al. Urinary Dickkopf-3 and Contrast-Associated Kidney Damage [J]. J Am Coll Cardiol, 2021, 77(21):2667-2676.

[10] Wu O, Hang J, Zhang XY, et al. The value of adiponectin-resistin (AR) index in newly diagnosed obesity hypertension: a case control study among Chinese adult [J]. Clin Exp Hypertens, 2022, 44(1):40-45.

[11] Sun XY, Ma RL, He J, et al. Updating Framingham CVD risk score using waist circumference and estimated cardiopulmonary function: a cohort study based on a southern Xinjiang population [J]. BMC Public Health, 2022, 22(1):1715.

[12] Li Y, Guo X, Liang S, et al. Endothelial function, arterial stiffness and Framingham risk score in chronic kidney disease: A prospective observational cohort study [J]. Hypertens Res, 2023, 46(4):868-878.

[13] Yu Q, Ning Y, Wang A, et al. Deep learning-assisted diagnosis of benign and malignant parotid tumors based on contrast-enhanced CT: a multicenter study [J]. Eur Radiol, 2023, 33(9):6054-6065.

[14] Zhang X, Hu W, Wang Y, et al. Plasma metabolomic profiles of dementia: a prospective study of 110655 participants in the UK Biobank [J]. BMC Med, 2022, 20(1):252.

[15] Meeusen JW, Kasozi RN, Larson TS, et al. Clinical Impact of the Refit CKD-EPI 2021 Creatinine-Based eGFR Equation [J]. Clin Chem, 2022, 68(4):534-539.

[16] Khan SS, Post WS, Guo X, et al. Coronary Artery Calcium Score and Polygenic Risk Score for the Prediction of Coronary Heart Disease Events [J]. JAMA, 2023, 329(20):1768-1777.

[17] Leening MJG, Steyerberg EW. Fibrosis and mortality in patients with dilated cardiomyopathy [J]. JAMA, 2013, 309:2547-2548.

（责任编辑：张　悦）