SPSS：详解临床预测模型的区分度和校准度

作者：龚志忠 来源：医咖会

本文原地址：https://www.iikx.com/news/statistics/11347.html

SPSS：详解临床预测模型的区分度和校准度
。基于风险预测模型的预后研究一直以来都是研究者关注的热点，各种各样的预测模型质量参差不齐，常常让人眼花缭乱，那么如何去评价一个模型的好坏，或者说当你构建出一个疾病风险预测模型后，它到底靠不靠谱，值不值得去推广和使用呢?这是一个我们需要去好好考量的问题。

一个好的疾病风险预测模型，它不只是简单的因变量和自变量的数学组合，它背后的实际临床意义才是我们所要把握的重点，这
就要求预测模型不仅要有很好的**区分度(Discrimination)**
，同时还要具备良好的**校准度(Calibration)**。

Discrimination和Calibration是我们在评价预测模型时最常用到的一对指标，但是2015年Circ Cardiovasc Qual Outcomes杂志(影响因子：4.5)上发表的一项关注心血管疾病预测模型的系统综述发现，63%的研究报告了预测模型的Discrimination信息，但仅36%的研究报告了Calibration信息，使得预测模型的质量成为研究泛滥的重灾区。

本文详细介绍一下临床预测模型的区分度和校准度这两个重要指标，尤其是常常被人忽略的Calibration。

区分度(Discrimination).

介绍Calibration之前，我们先简单介绍一下Discrimination。顾名思义，一个好的疾病风险预测模型，它能够把未来发病风险高、低不同的人群正确地区分开来，预测模型通过设置一定的风险界值，高于界值判断为发病，低于界值则判断为不发病，从而正确区分个体是否会发生结局事件，这就是**预测模型的区分度(Discrimination)**。

评价预测模型区分能力的指标，最常用的就是大家非常熟悉的ROC曲线下面积(AUC)，也叫C统计量(C-statistics)。
AUC越大，说明预测模型的判别区分能力越好
。一般AUC<0.6认为区分度较差，0.6-0.75认为模型有一定的区分能力，>0.75认为区分能力较好
。

校准度(Calibration).

预测模型的校准度(Calibration)，是

评价一个疾病风险模型预测未来某个个体发生结局事件概率准确性的重要指标，它反映了模型预测风险与实际发生风险的一致程度
，所以也可以称作为一致性
。校准度好，提示预测模型的准确性高，校准度差，则模型有可能高估或低估疾病的发生风险。

在实际的应用中，通常用Hosmer-Lemeshow good of fit test(拟合优度检验)来评价预测模型的校准度。Hosmer-Lemeshow检验的基本思路如下：

1. 首先根据预测模型来计算每个个体未来发生结局事件的预测概率;

2. 根据预测概率从小到大进行排序，并按照十分位等分成10组;

3. 分别计算各组的实际观测数和模型预测数，其中模型预测数，即每个人的预测概率*人数，再求总和，这里人数即为1，最后总和就相当于每个个体预测概率的直接加和;

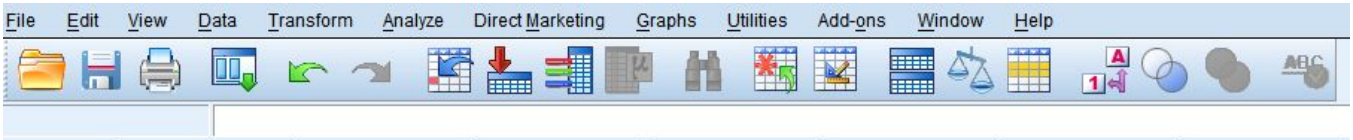
4. 根据每组实际观测数和模型预测数计算卡方值(自由度=8)，再根据卡方分布得到对应的P值。

若所得的统计量卡方值越小，对应的P值越大，则提示预测模型的校准度越好。若检验结果显示有统计学显著性(P<0.05)，则表明模型预测值和实际观测值之间存在一定的差异，模型校准度差。

区分度和校准度的SPSS操作:

一、建立数据库.

某研究人员拟建立一个关于冠心病患者支架介入术后再次发生MACE事件(Major Adverse Cardiovascular Events，主要心血管不良事件)的风险预测模型，并对该风险模型的预测能力进行评价。数据库格式如下图所示。



	Event	Gender	Age	SBP	Smoking	LDL	SYNTAX
1	0	1	82	140	0	1.53	20.0
2	0	1	79	140	0	2.05	17.0
3	0	1	79	140	0	2.82	24.0
4	1	1	78	130	1	3.92	16.5
5	0	1	78	137	0	3.21	32.0
6	1	1	77	140	1	1.89	44.0
7	0	1	77	140	0	1.36	11.0
8	0	1	77	130	0	1.65	14.0
9	1	1	76	137	1	3.00	21.5
10	1	1	76	145	0	3.02	9.0

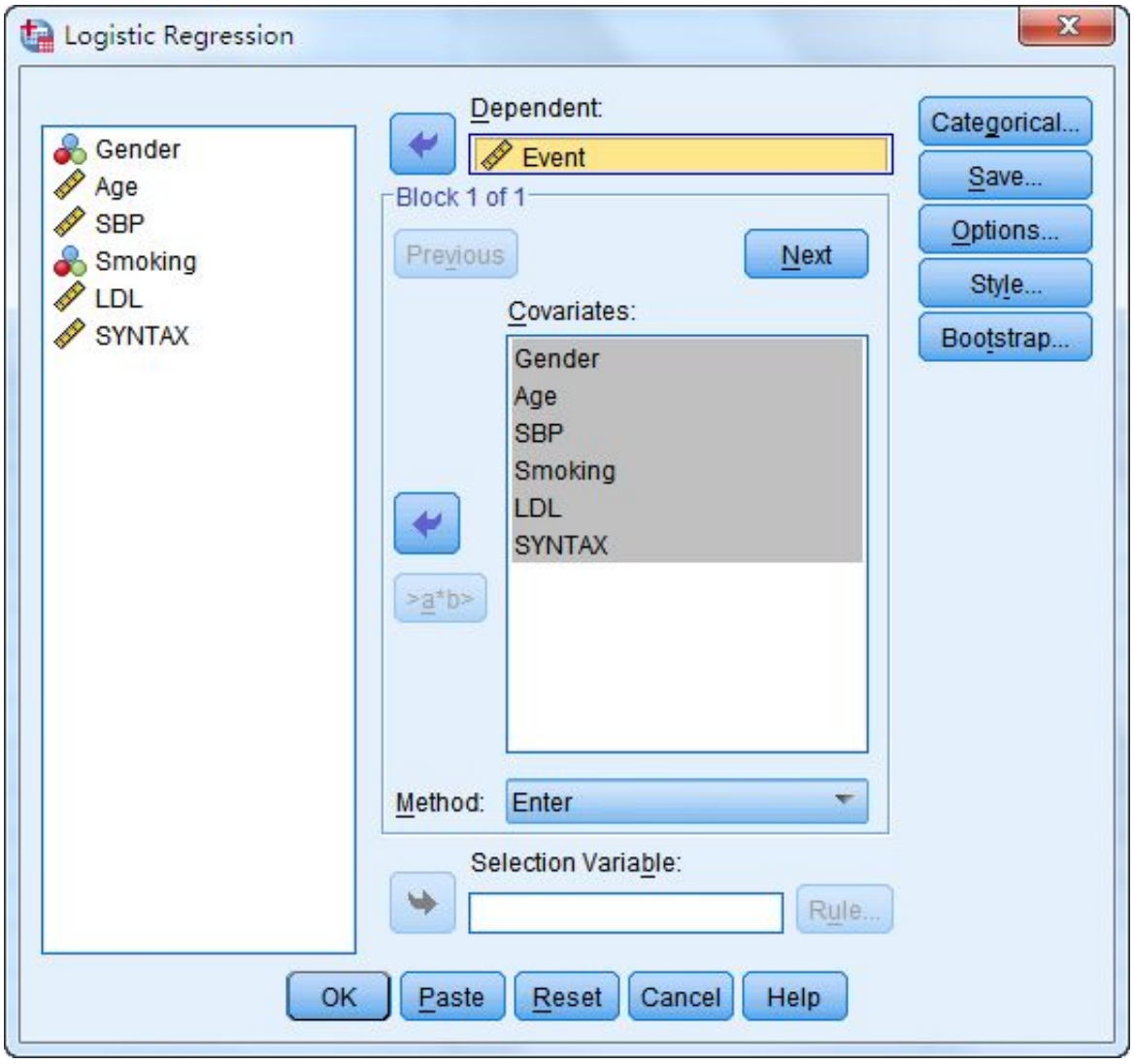
其中因变量(结局事件)为Event，自变量(影响因素)为性别(Gender)、年龄(Age)、收缩压(SBP)、吸烟(Smoking)、低密度脂蛋白胆固醇(LDL)及冠脉病变Syntax评分(SYNTAX)。

二、构建预测模型.

本研究利用Logistic回归构建预测模型(若研究为含有时间变量的生存数据，则可采用Cox回归模型)。Logistic回归的操作步骤对大家来说应该早就是小case了，操作方法如下：

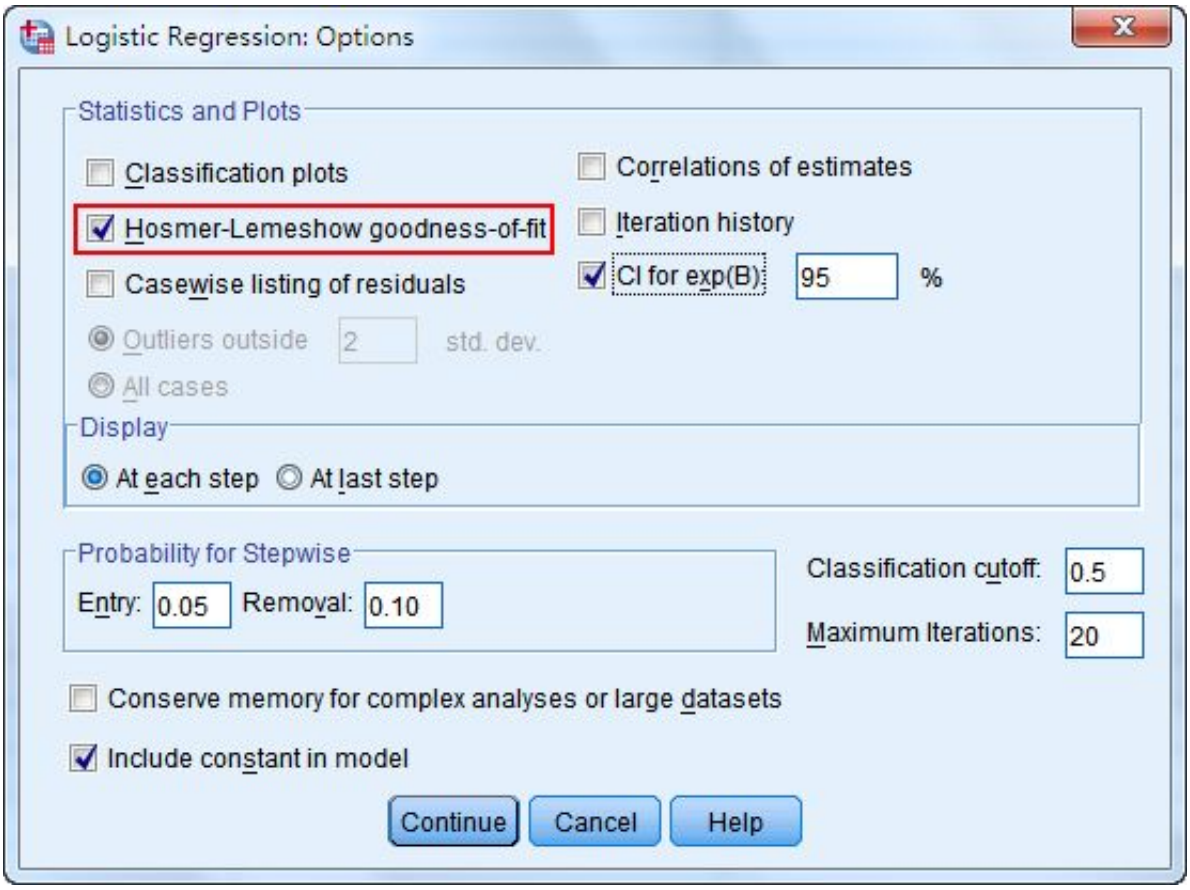
1. Analyze → Regression → Binary Logistic Regression

2. 将因变量Event选入Dependent框中，将各个自变量选入Covariates框中



3. 点击Save，在Predicted Values下勾选Probabilities，目的是为了在数据库中新生成一个概率值，用于绘制ROC曲线和校准曲线图。

4. 点击Options，勾选Hosmer-Lemeshow goodness-of-fit，用于输出Hosmer-Lemeshow拟合优度检验的结果。



三、Logistic回归结果.

Variable in the Equation中输出了每个影响因素的回归系数(β)、OR值、95% CI以及P值等信息。回归方程如下：

logit (p)= -8.713 - 0.899*Gender + 0.05*Age + 0.021*SBP + 0.912*Smoking + 0.438*LDL + 0.07*SYNTAX

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Gender	-.899	.360	6.222	1	.013	.407	.201	.825
	Age	.050	.016	9.755	1	.002	1.051	1.019	1.085
	SBP	.021	.009	5.788	1	.016	1.021	1.004	1.038
	Smoking	.912	.297	9.455	1	.002	2.490	1.392	4.453
	LDL	.438	.177	6.134	1	.013	1.549	1.096	2.190
	SYNTAX	.070	.023	9.661	1	.002	1.073	1.026	1.121
	Constant	-8.713	1.607	29.403	1	.000	.000		

四、模型区分度(Discrimination).

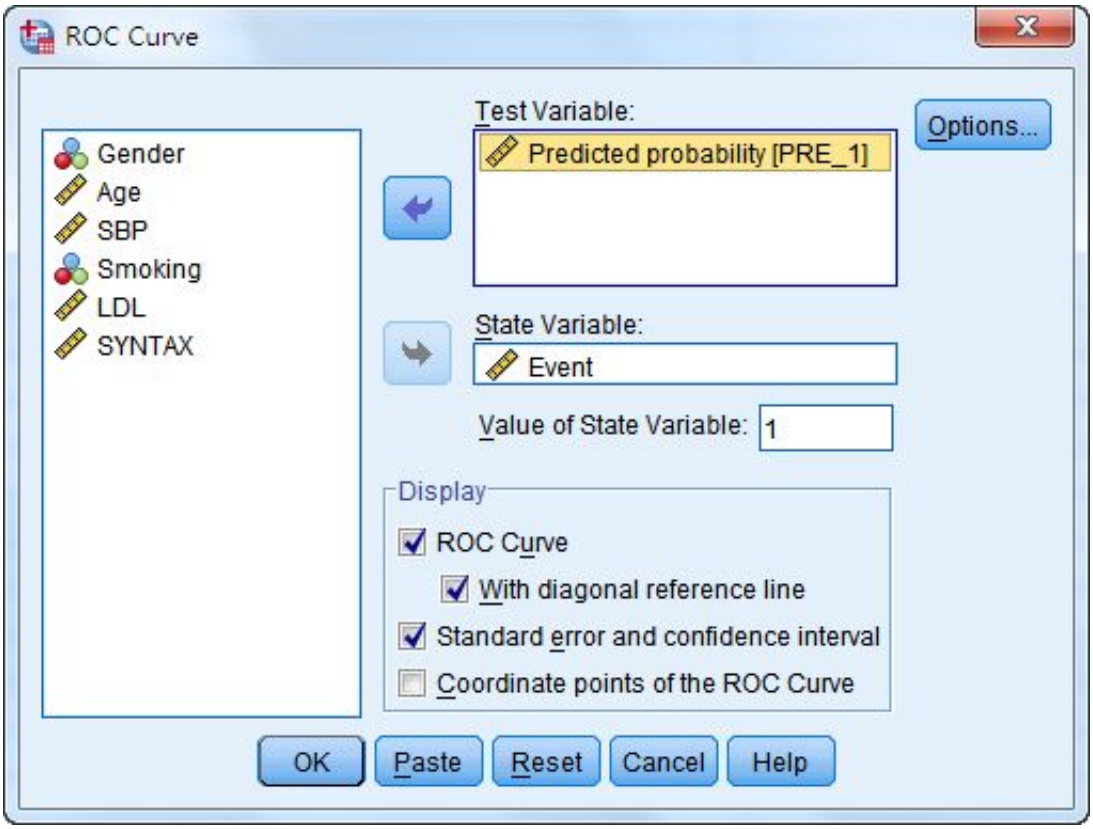
通过绘制ROC曲线，计算AUC，即C统计量来评价模型的判别区分能力。具体操作步骤为：

1. Analyze → ROC Curve

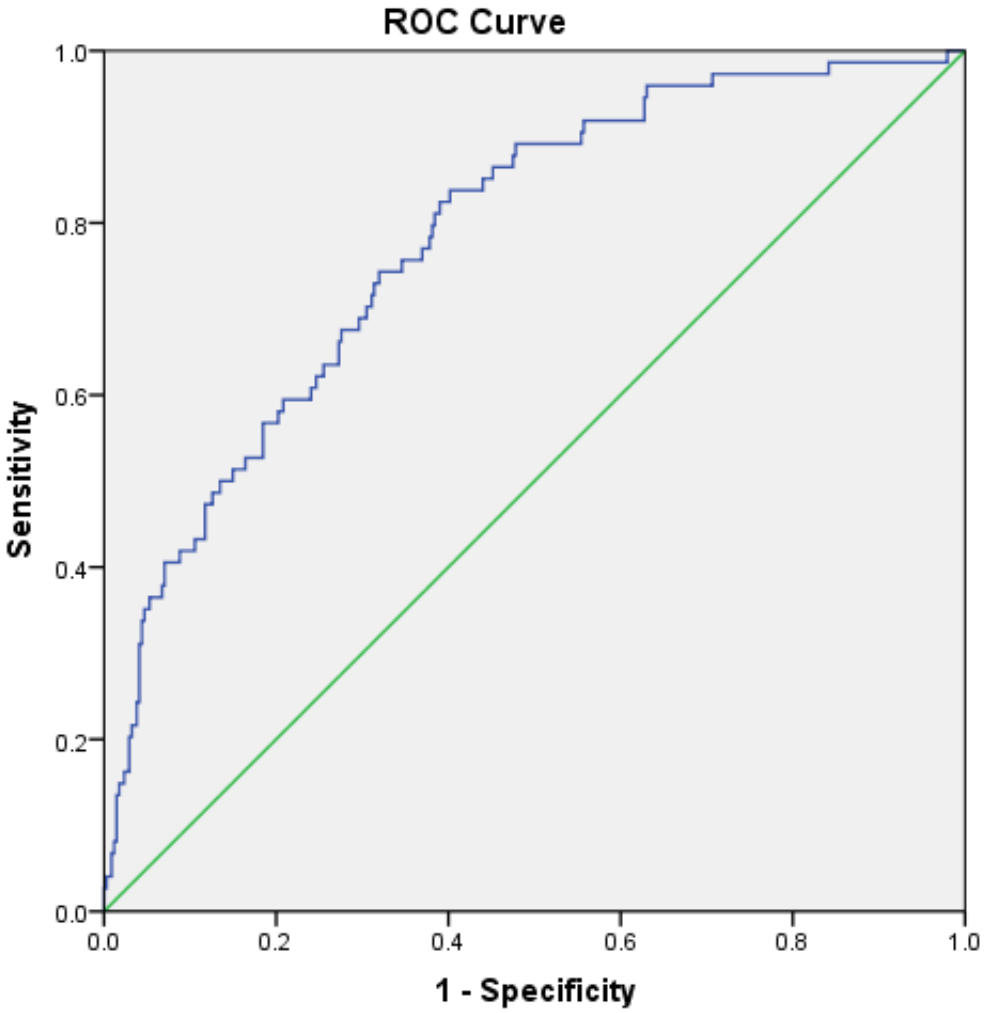
2. 将新生成的预测概率值PRE_1作为检验变量Test Variable，将Event作为状态变量State Variable，并设定Value of State Variable为1

3. 勾选ROC Curve用于绘制ROC曲线

勾选Standard error and confidence interval用于输出AUC及其标准误和95%可信区间。



预测模型ROC曲线如下图所示，曲线下面积AUC为0.782>0.75，95% CI为0.726-0.838，提示该预测模型的区分能力较好。



Area Under the Curve

Test Result Variable(s): Predicted probability

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.782	.029	.000	.726	.838

五、模型校准度(Calibration).

通过Hosmer-Lemeshow拟合优度检验来评价预测模型的校准能力。结果显示，Hosmer-Lemeshow $\chi 2=4.864$，P=0.772>0.05，提示模型预测值与实际观测值之间的差异没有统计学显著性，预测模型有较好的校准能力。

同时SPSS还输出了Hosmer-Lemeshow检验列联表，表中将每个研究对象的预测概率从小到大进行排序，并按照十分位分成10组，分别列出了每一组实际观测值(Observed)和模型预测值(Expected)，从而可以在每一个分组下进行直观的比较，来帮助判断模型的校准能力。

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	4.864	8	.772

Contingency Table for Hosmer and Lemeshow Test

		Event = 0		Event = 1		
		Observed	Expected	Observed	Expected	Total
Step 1	1	41	41.003	1	.997	42
	2	41	40.169	1	1.831	42
	3	41	39.331	1	2.669	42
	4	37	38.399	5	3.601	42
	5	38	37.334	4	4.666	42
	6	34	35.818	8	6.182	42
	7	32	33.830	10	8.170	42
	8	34	31.503	8	10.497	42
	9	29	27.503	13	14.497	42
	10	14	16.111	23	20.889	37

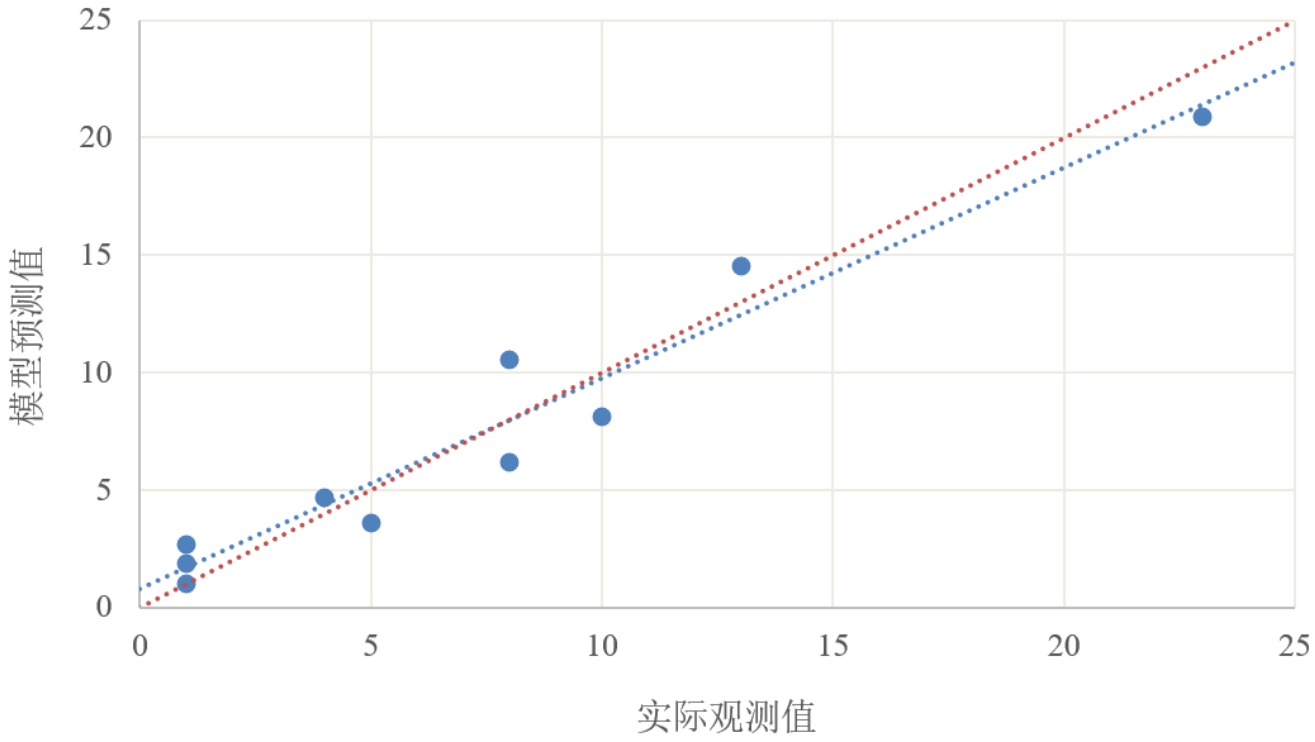
六、模型校准图形(Calibration Plot).

既然在评价预测模型区分度的时候，结果可以通过绘制ROC曲线进行可视化，那么对于预测模型的校准度，我们也同样可以绘制校准图使结果可视化。

我们在文献中常常可以看到，校准图的绘制一般有三种形式，大家可以利用上面SPSS输出的Hosmer-Lemeshow检验列联表的结果，将其复制到Excel中(以下图形均以Excel 2013版为例)，跟着小咖一起来绘制校准图形。

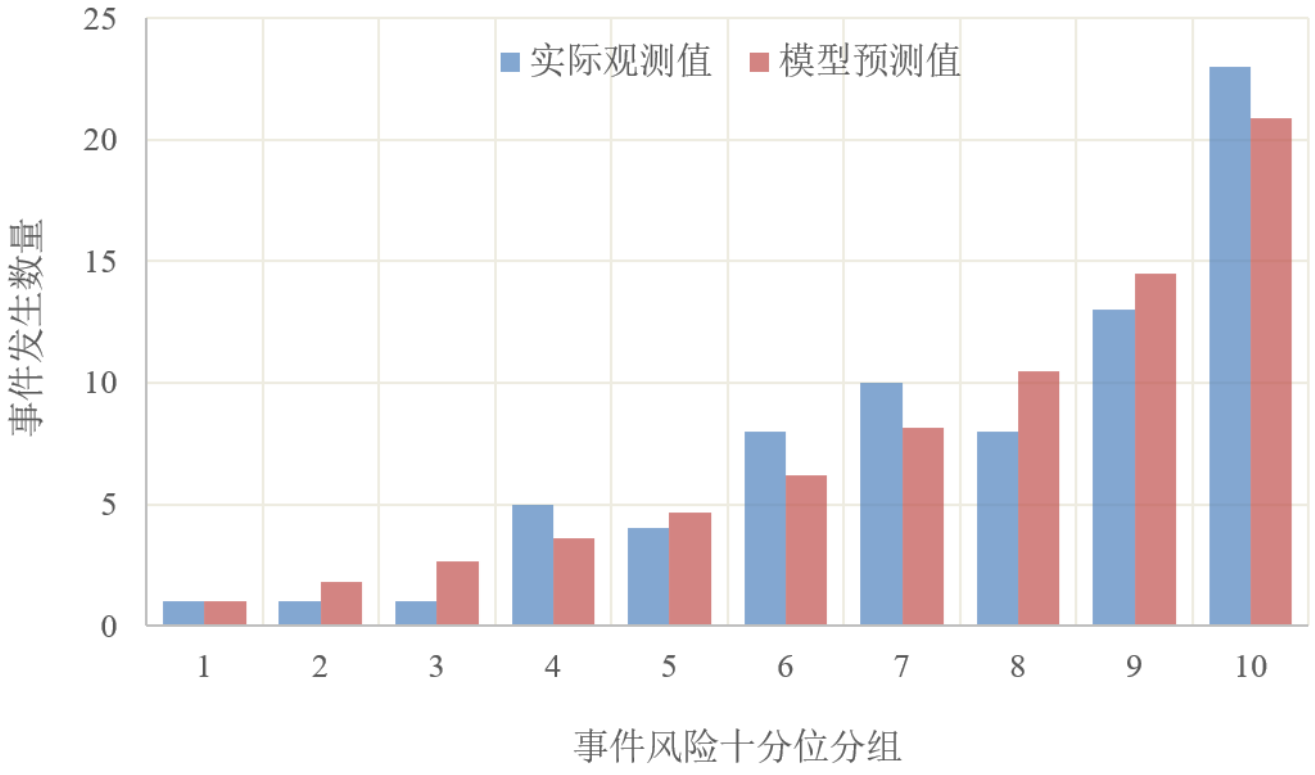
1. 散点图.

根据实际观测值(Observed)和模型预测值(Expected)绘制散点图，并拟合线性趋势线，即可得到校准曲线，如下图所示的蓝线。而红线为标准曲线(y=x)，表示预测数和实际观测数完全一样。若蓝色的校准曲线和红色的标准曲线越接近，则提示模型的校准能力越好。



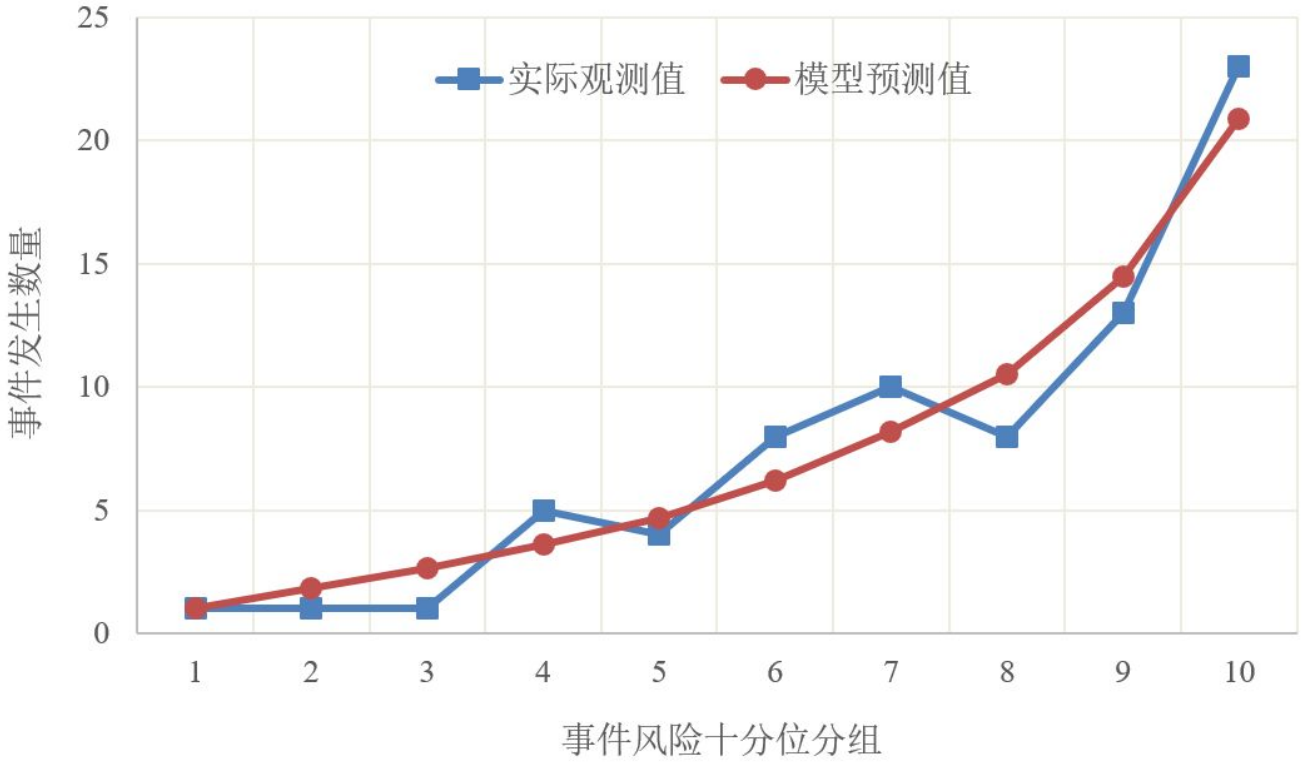
2. 条形图.

将每个研究对象的预测概率从小到大进行排序，并按照十分位分成10组，以条图的形式来表示每组实际观测值和模型预测值的大小，这样能够更加直观的展示在每一组内，实际观测值和模型预测值之间的差别，以此来帮助判断模型更为准确的预测区间。



3. 线图.

线图的表达方式和条形图类似，同样也是按照预测概率的十分位分成10组，以坐标点的形式来表示每组实际观测值和模型预测值的大小，并用平滑的线段依次连接起来。它不仅可以直观的展示每一组内实际观测值和模型预测值之间的差别，同时也能从整体上来判断模型的校准能力。模型预测曲线与实际观测曲线越接近，则可提示模型的校准能力越好。



总结.

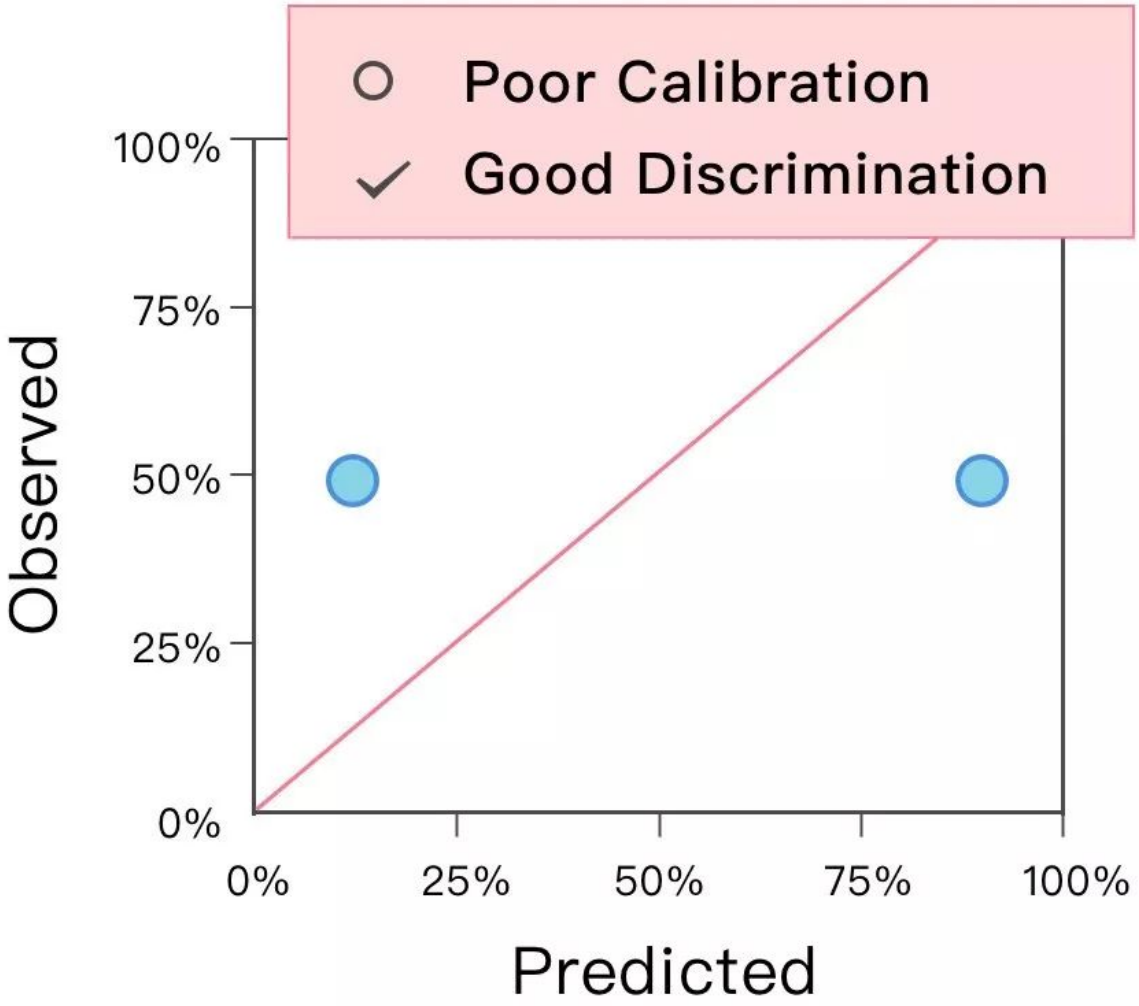
Discrimination和Calibration是评价预测模型效能的两个重要指标，但比较容易混淆，最后再和大家总结一下：

1. Discrimination区分度，就是在模型的预测值中，看是否能够找到一个截点，使得把患者和非患者正确区分开来。如果区分的越开，且与实际情况越吻合，则提示模型的区分度越好。

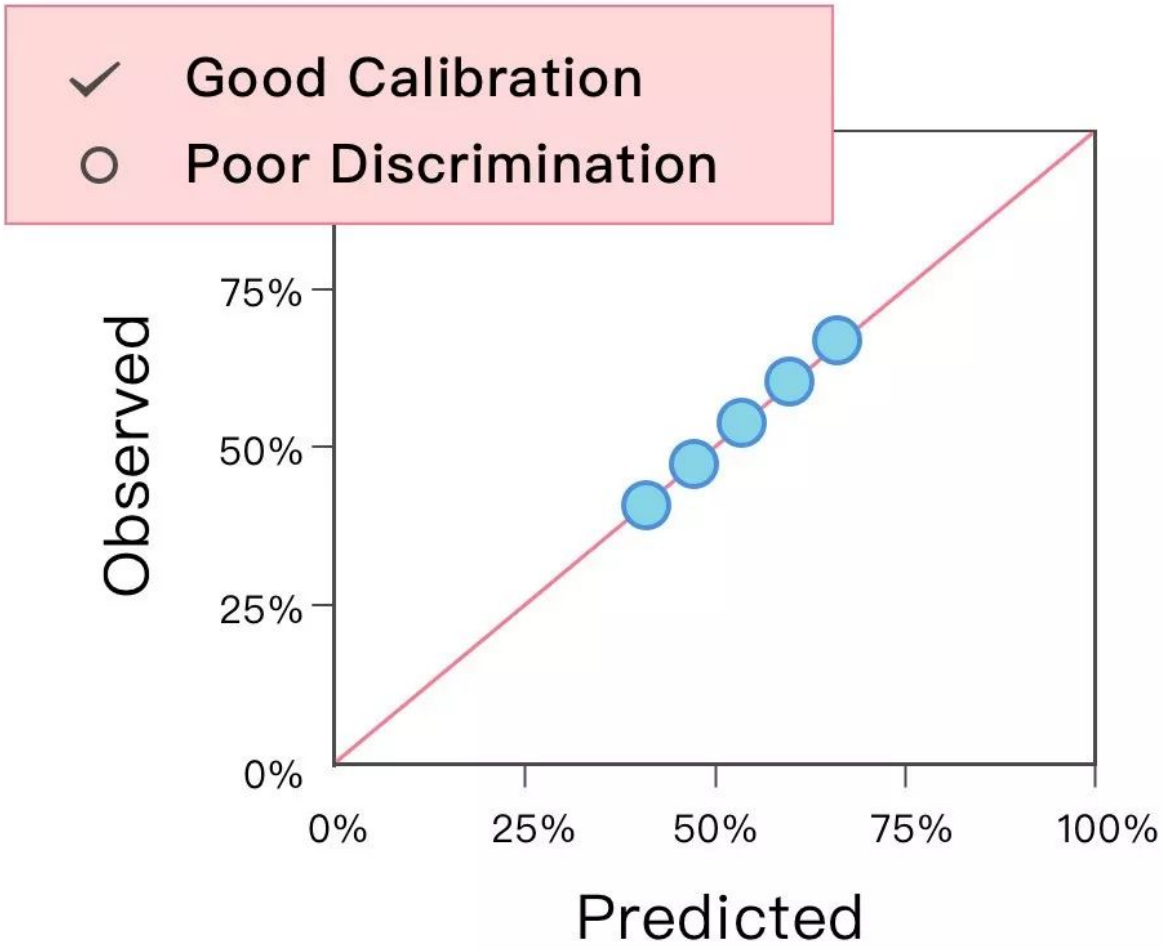
2. Calibration校准度，就是评价模型预测值的大小和结局事件发生概率的大小是否一致。如果模型的预测值与结局实际发生概率越接近，则提示模型的校准度就越好。

3. 风险预测模型的Discrimination和Calibration并不一定都是同方向的。

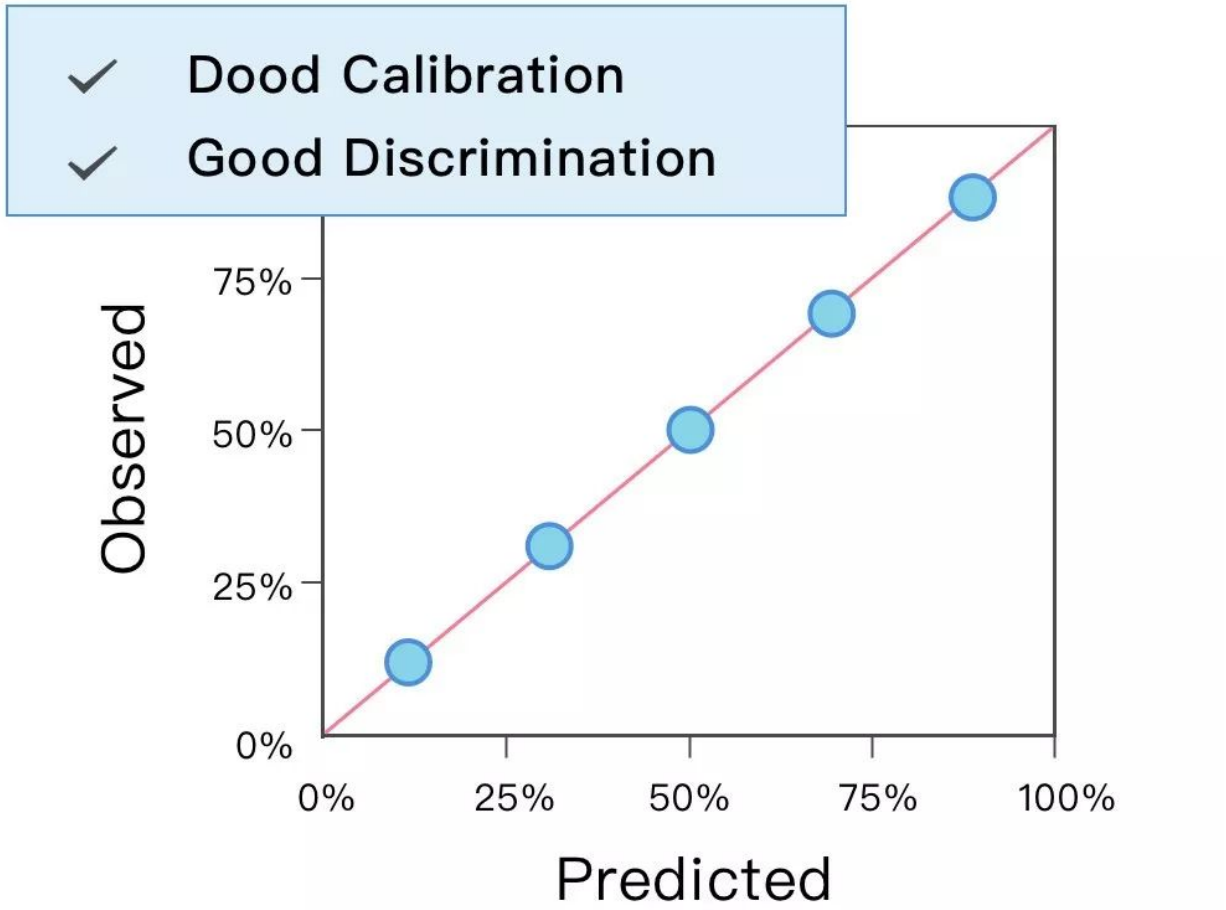
如图A，模型的Discrimination很好，能够根据发病风险将不同的研究对象明显的区分开来，但是Calibration较差，预测值偏离校准曲线很远，与实际情况不符。


Poor Calibration
Good Discrimination
100%
75%
50%
25%
0%
Observed
0%
25%
50%
75%
100%
Predicted

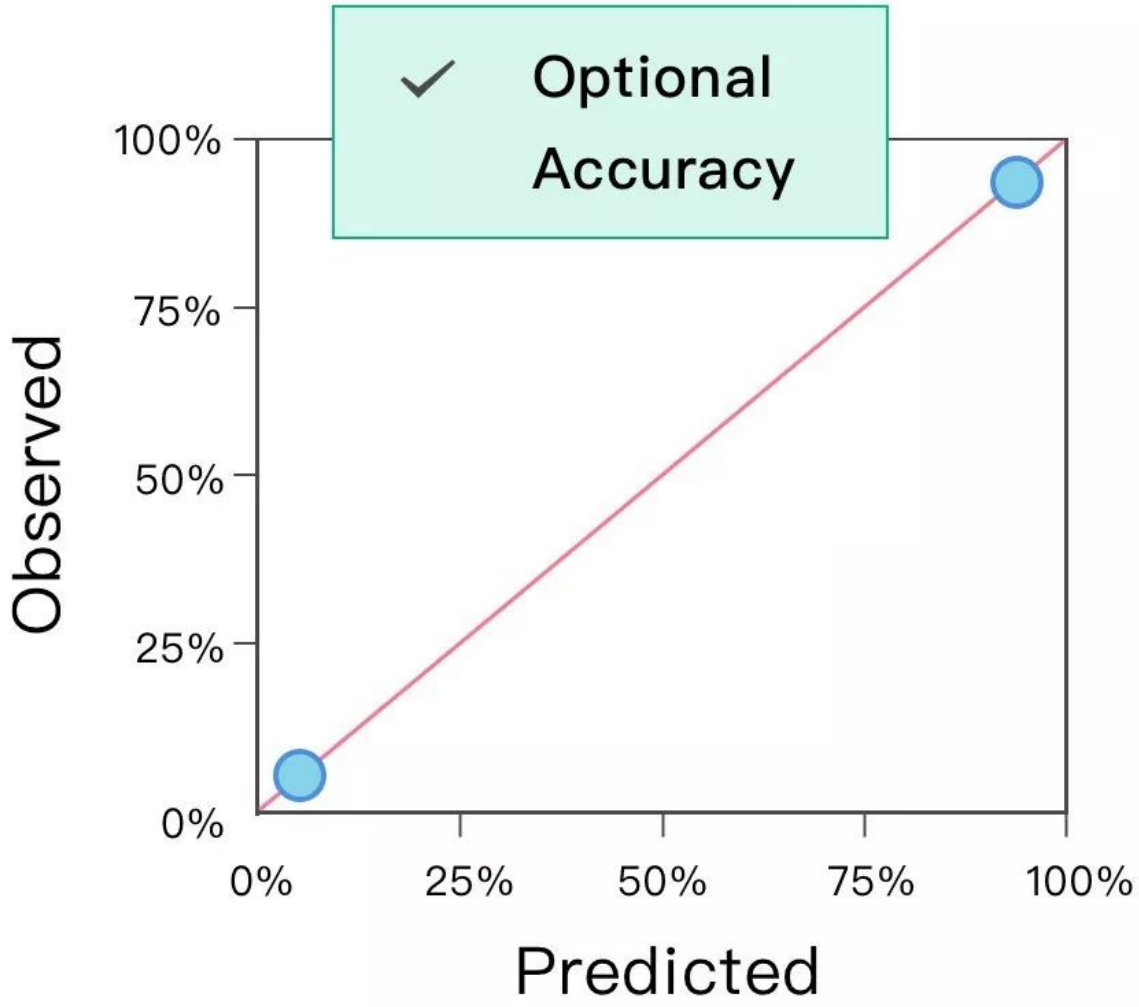
(图A)



(图B)，模型的Calibration很好，预测值都集中在校准曲线上，预测较为准确，但是Discrimination较差，研究对象的患病风险都比较接近，无法将其明显的区分开来。



(图C)，模型的Discrimination和Calibration都很好，不仅能够把不同风险的患者明显的区分开来，而且预测值都集中在校准曲线上，预测结果较为准确。



(图D)，是最为理想的模型，能够准确预测研究对象是否患者，发病风险为0或100%。

4. 对于一个疾病预测模型，在利用Discrimination和Calibration进行评价时，我们首先需要考虑的是模型的区分能力Discrimination，如果模型的区分度较差，不能正确的将不同风险的人群区分开来，那么它就不是一个合格的预测模型，失去了临床的应用价值，再继续评价Calibration也没有太大的意义了。

所以，如果你对自己建立的疾病风险预测模型有足够的信心，那么不妨也计算一下模型的Discrimination和Calibration，相信一定会得到更多同行的认可。

更多 统计方法 请访问 https://www.iikx.com/news/statistics/