

预警预测模型构建

张效嘉

军事医学科学院

一、预测患者一年内不良事件发生概率

数据集：B.sav

因变量：MACE（不良事件）：0—未发生，1—发生

自变量：根据郑医生的建议，初始自变量集合如下：

4 个定量变量：Age, LVEF, Hb_Admission, Crea_Admission

22 个二值变量：Sex DM HTA DLP PriorAMI PriorPCI PriorCABG

Hx_Stroke Hx_CHF CKD Malignancy STEMI LVEF Hb_Admission

Crea_Admission Multivessel DES Thrombolysis Revasc_Complete Clopi

OAC BB ACEI_ARB Statin PPI

模型构建过程如下：

1. 自变量筛选

采用方法：逐步判别分析，其中纳入标准 $\alpha_{\text{纳入}} = 0.10$ ，剔除标准 $\alpha_{\text{剔除}} = 0.15$

结果：筛选得到的自变量如下：

4 个定量变量：Age, LVEF, Hb_Admission, Crea_Admission

12 个二值变量：Sex Hx_CHF Revasc_Complete BB PriorPCI HTA

Malignancy PPI OAC DM DES STEMI

2. 一年内不良事件发生概率预测

采用方法：Bayes 判别分析与 Bayes 公式判别分析结合后的改进算法，具体如下（先验概率设为构成比，发生先验概率为 q_1 ，不发生先验概率为 q_2 ）：

1) 以 MACE 为因变量，以筛选后得到的 4 个定量变量为自变量，采用交叉验证 Bayes 判别分析计算各患者 1 年内不良事件的发生后验概率 p_1 和不发生后验概率 $p_2(p_1+p_2=1)$

2) 以 MACE 为因变量，以筛选后得到的 12 个二值变量为自变量，采用 Bayes 公式判别分析计算各患者 1 年内不良事件的发生后验概率 p_3 和不发生后验概率 $p_4(p_3+p_4=1)$

3) 以下面公式计算各患者一年内不良事件发生概率：

$$P = \frac{p1 \cdot \frac{p3}{q1}}{\frac{p1 \cdot p3}{q1} + \frac{p2 \cdot p4}{q2}}$$

结果：各患者一年内不良事件发生概率（post1）详见附件。

3. 模型拟合效果评价

采用方法：ROC 曲线下面积，具体如下：

选取不同概率阈值。在特定概率水平下，计算灵敏度（从 MACE=1 划分到 MACE=1 中的频率）和特异度（从 MACE=0 划分到 MACE=0 中的频率）。以灵敏度为纵坐标、1-特异度为横坐标绘制 ROC 曲线，并计算曲线下面积。

二、患者使用某种药物或治疗措施的增益

数据集：B.sav

因变量：MACE（不良事件）：0—未发生，1—发生

自变量：根据郑医生的建议，初始自变量集合如下：

4 个定量变量：Age, LVEF, Hb_Admission, Crea_Admission

22 个二值变量：Sex DM HTA DLP PriorAMI PriorPCI PriorCABG
Hx_Stroke Hx_CHF CKD Malignancy STEMI LVEF Hb_Admission
Crea_Admission Multivessel DES Thrombolysis Revasc_Complete Clopi
OAC BB ACEI_ARB Statin PPI

模型构建过程如下：

1. 自变量筛选

采用方法：COX 比例风险回归的逐步选择法

首先固定 LVEF STEMI OAC BB ACEI_ARB Statin PPI 这 7 个自变量选入模型，再利用逐步选择法对剩下的变量进行筛选，其中纳入标准 $\alpha_{\text{纳入}} = 0.10$ ，剔除标准 $\alpha_{\text{剔除}} = 0.15$ 。希望临床专业人员提出有价值的分层因素以及变量间的交互作用（重要）。

结果：筛选得到的 16 个自变量如下：

LVEF STEMI OAC BB ACEI_ARB Statin PPI Age Sex DM HTA
PriorPCI Malignancy Hb_Admission DES Revasc_Complete

2. 患者使用某种药物或治疗措施的增益

采用方法：COX 比例风险回归模型

结果：

参数	参数估计	标准误差	卡方	Pr > 卡方	危险比	95% 危险比置信限	
LVEF	-0.02227	0.00473	22.1303	<.0001	0.978	0.969	0.987
STEMI	-0.24926	0.12395	4.0440	0.0443	0.779	0.611	0.994
OAC	0.36731	0.19193	3.6624	0.0557	1.444	0.991	2.103
BB	-0.39014	0.12795	9.2971	0.0023	0.677	0.527	0.870
ACEI_ARB	-0.00778	0.13973	0.0031	0.9556	0.992	0.755	1.305
Statin	0.01039	0.21575	0.0023	0.9616	1.010	0.662	1.542
PPI	0.27790	0.12966	4.5942	0.0321	1.320	1.024	1.702
Age	0.02739	0.00595	21.1625	<.0001	1.028	1.016	1.040
Sex	-0.43173	0.14432	8.9485	0.0028	0.649	0.489	0.862
DM	0.21043	0.12414	2.8733	0.0901	1.234	0.968	1.574
HTA	0.27686	0.13354	4.2985	0.0381	1.319	1.015	1.714
PriorPCI	0.49208	0.16936	8.4420	0.0037	1.636	1.174	2.280
Malignancy	0.38043	0.17580	4.6830	0.0305	1.463	1.037	2.065
Hb_Admission	-0.17526	0.03439	25.9699	<.0001	0.839	0.785	0.898
DES	-0.27809	0.13132	4.4843	0.0342	0.757	0.585	0.980
Revasc_Complete	-0.40331	0.11675	11.9330	0.0006	0.668	0.531	0.840

分析：

其中，参数估计的值是风险函数中各自变量前的系数

（某药物）危险比是指用药与不用药的相对风险，95%危险比置信限是指这种相对风险的 95%置信区间。

另外，预后指数(PI)和生存率(S)如有需要，请见附件。

注意（非常重要）：如果以上表格结果有违医学常理，说明一些自变量间的交互作用未被引入。此时，需要从专业角度引入变量间的交互作用（如 drug1*drug2）