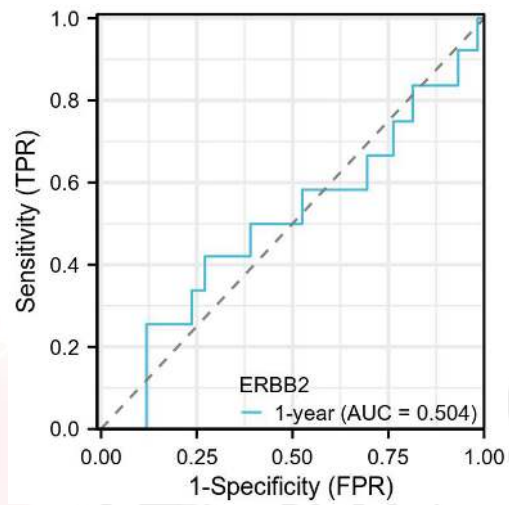


临床意义 - 时间依赖 ROC - 云



网址: <https://www.xiantao.love>



更新时间: 2023.03.03

目录

基本概念	3
应用场景	3
分析流程	4
结果解读	6
云端数据	8
参数说明	9
特殊参数	9
预后参数	10
预测时间	11
置信区间	12
线	13
点	14
曲线下面积	15
标题文本	16
图注	17
风格	18
图片	19
结果说明	20
主要结果：单个预测变量单个预测时间点	20
主要结果：单个预测变量多个预测时间点	20
主要结果：多个预测变量多个预测时间点	21
补充结果：单个预测变量多个预测时间点的情况	22
补充结果：多个预测变量多个预测时间点的情况（并且计算置信区间的前提下）	24
方法学	26
如何引用	27
常见问题	28

基本概念

- 受试者工作特征曲线 (receiver operating characteristic curve, ROC) : 一种坐标图式的分析工具, 可用于选择最佳的模型、舍弃次佳的模型或者在同一模型中设定最佳阈值。比如, 一些新标志物的预测模型会用 ROC 曲线来评价其性能。
- 时间依赖 ROC: ROC 曲线在生存资料类型的模型中的一种特殊的应用, 对于给定的时间点, 如果一个样本在时间 0 和时间 t 之间发病, 那么其会被划分到病例组, 否则则划分到对照组。对于每一个自变量不同的阈值划分分组最终就能得到每个阈值下的敏感度、特异度, 获得最佳阈值并且评估生存资料类型模型的好坏。

应用场景

ROC 曲线主要用于 连续变量来预测带有生存资料的结局事件, 找到连续变量最佳的 cut-off 值; 或者用于评估变量对于带有生存资料的结局的预测效能

此处的 ROC 模块主要用于分析 分子表达与时间相关的预测结局的预测效能!

分析流程

云端数据 → ROC 相关分析 → ROC 曲线可视化

➤ 云端数据：不同平台的云端数据集的分子可能会有不同!

- 通过特殊参数[分子]选择云端数据中需要进行分析以及可视化的分子/变量

特殊参数

分子

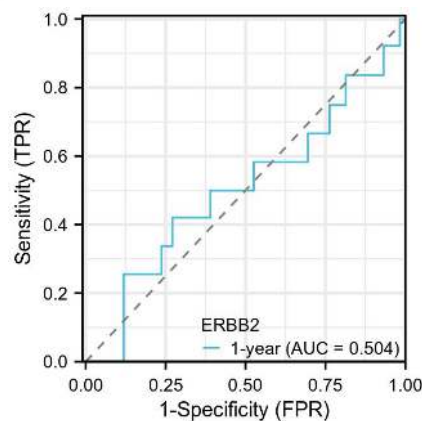
- 通过主要参数[预后参数-预后类型]选择不同的预后类型进行分析
 - ◆ 预后类型：可选不同的预后类型。不同的数据集之间的预后类型可能不一样!
 - OS[Overall Survival]: 总体生存期，指结局指标是死亡时间，这个死亡是任何原因导致的死亡都算进去，只关心是否死亡，不关心因为何种原因死亡
 - DSS[Disease Specific Survival]: 无病生存期，指经过治疗后未发现肿瘤，结局指标为疾病复发或死亡，同样不需要关心死亡原因。这一指标是临床获益的重要反映，随访时间可以缩短，因为增加了疾病复发这一节点。没有复发或 没有死亡同样可以反映临床获益。这里也涉及到无疾病复发的一个定义，因此在临床资料纳入上比较困难
 - PFI[Progress Free Interval]: 无进展间隔，从初次治疗的随机分组日期到疾病复发时间。（具体可以参考对应的引文）

➤ ROC 相关分析：

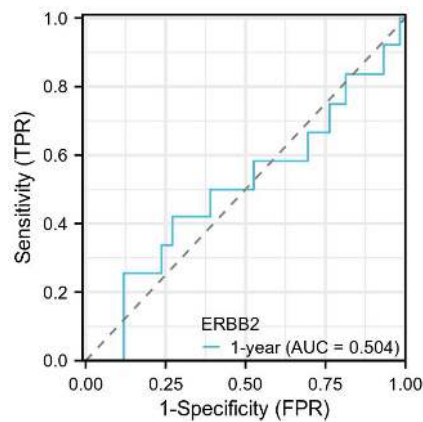
- 统计描述：在预测的某个时间点，统计各分子的生存情况
- ROC 结果：在预测的某个时间点，计算各分子的最佳阈值（cut-off 值）、敏感度、特异度等
- AUC 信息：在预测的某个时间点，计算各分子 ROC 曲线下面积（AUC）、累计生存率、累计发病率等
-

➤ 将分析过程中得到的敏感度、特异度数据进行转换，得到可视化需要的数据，并进行可视化

- 敏感度 = 真阳性率(TPR)
- 特异度 = 1 - 假阳性率(FPR)
- 结果如下：



结果解读



- 测试横坐标代表 1-Specificity (FPR)，即“1-特异度”，代表假阳性率。FPR 越大,预测假阳性率越高
- 纵坐标代表 Sensitivity (TRP)，即灵敏度，代表真阳性率
- ROC 曲线的是反映敏感性和特异性连续变量的综合指标,通过构图法体现敏感性和特异性的相互关系。可以理解为将连续变量设定出多个不同的临界值，从而计算出一系列敏感性和特异性，再以敏感性为纵坐标
- 当一个分子(危险因素)的值是促进事件发生的趋势时,该分子的 AUC 会 >0.5 ，此时面积越大(AUC 值越接近于 1)说明预后效能越好
 - AUC 在 0.5~0.7 时，预后效能较低
 - AUC 在 0.7~0.9 时，有一定预后效能
 - AUC 在 0.9 以上时，预后效能较好
- 当一个分子(保护因素)的值与事件发生的趋势相反时，则这个分子的 AUC 会 <0.5 ，此时面积越小(AUC 值越接近于 0)说明预后效能越好

补充：

- timeROC 包固定了（无法修改）在一个时间点内发生事件的作为病例组，所以当一个分组数值的趋势与预后结局趋势相反（类似于保护因素 $HR < 1$ ）时，则会出现 $AUC < 0.5$
- 如果某个时间点对应的曲线下面积为 NA，表示该时间点下曲线下面积不存在
- 如果某个时间点的事件数为 0，则该时间点无法进行相关的 ROC 分析和可视化

云端数据

提供预清洗好的云端数据, 不同平台的云端数据集的分子可能会有不同!

(该样本数据如下:)

数据参数 重置参数

云端数据 ⓘ 食管鳞癌 / TCGA / TCGA-ESCC / RNAseq / STAR / TPM @过滤:去除正常+去除无临床信息 @处理:log2(...)



参数说明

(说明: 标注了颜色的为常用参数。)

特殊参数

特殊参数

分子 ?

ERBB2|ENSG00000141736.14| ×

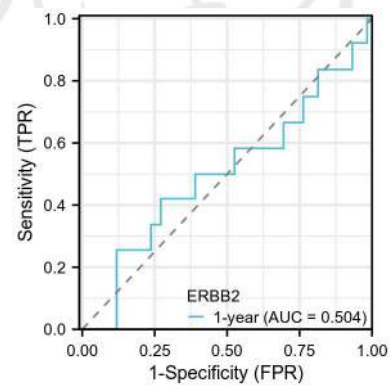
- 分子: 选择云端数据中需要进行分析以及可视化的分子/变量, 可以选择多个, 并且至少需要一个

- 可以输入关键词搜索分子
- 当选择了多个分子时, 只会分析当前预测时间中对应设置的最小的时间

特殊参数

分子 ?

ERBB2|ENSG00000141736.14| ×



特殊参数

分子 ?

ERBB2|ENSG00000141736.14| ×

TSPAN6|ENSG00000000003.15| ×

TNMD|ENSG00000000005.6| ×

1-year
ERBB2 (AUC = 0.504)
TSPAN6 (AUC = 0.499)
TNMD (AUC = 0.534)

9

预后参数

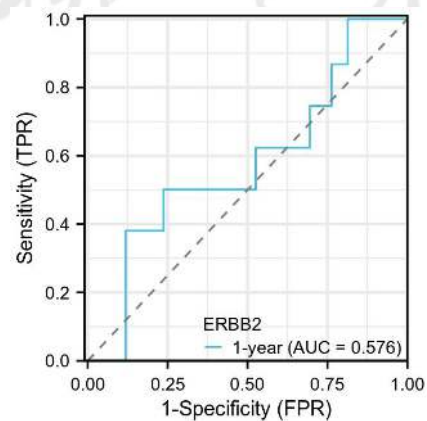
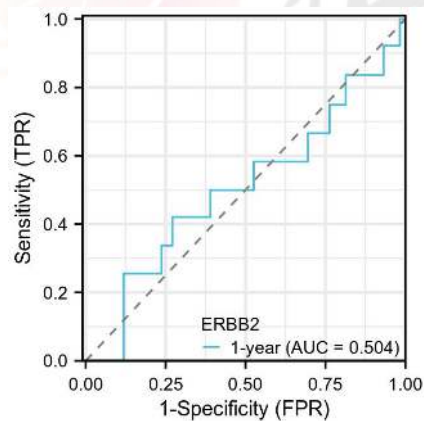
预后参数

预后类型 OS[Overall !

➤ 预后类型：可选不同的预后类型。不同的数据集之间的预后类型可能不一样！，可以选择：

- OS[Overall Survival] (默认)：总体生存期
- DSS[Disease Specific Survival]：无病生存期
- PFI[Progress Free Interval]：无进展间隔

➤ 如下：左侧为：OS[Overall Survival]，右侧为DSS[Disease Specific Survival]



预测时间

预测时间

时间1

1

时间2

请输入数字

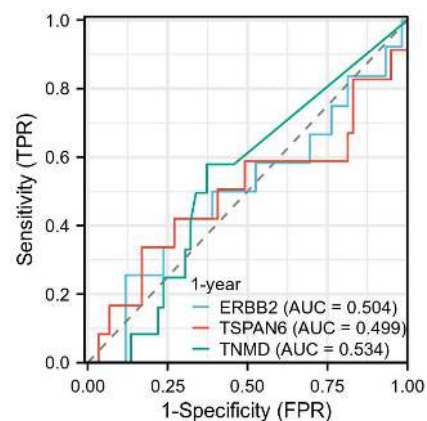
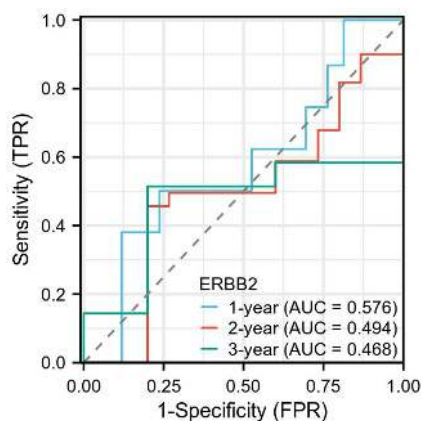
时间3

请输入数字

时间单位

年

- **时间 1**：预测的第一个时间点
- **时间 2**：预测的第二个时间点
- **时间 3**：预测的第三个时间点
- **时间单位**：可以选择上传数据预测时间列的单位，默认以年为单位，还可以选择月、天为单位
- 最多只能同时预测 3 个时间点（前提是只有一个预测分子的时候），当选择了多个分子时，只会分析当前预测时间中对应设置的最小的时间，如下（左为单个预测分子 3 个预测时间，右为多个预测分子 3 个预测时间）



置信区间

置信区间

是否计算

- 是否计算：可以选择是否需要进行 AUC（曲线下面积）置信区间相关计算、或在多个预测分子进行分析时可以计算其组间检验，对于数据量比较大的情况下，强制是不会进行进行置信区间的计算的

■ 默认为不计算

AUC信息表					
变量	时间(年)	天数	曲线下面积(AUC)	累计生存率	累计发病率
ERBB2	1	365	0.57633	0.88793	0.11207

· 通过timeROC包计算得到每个时间点对应的曲线下面积和累计生存率
· 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算对应的曲线下面积

- 还可以选择进行计算，表示会计算 AUC 置信区间，结果会在补充结果部分展示：

AUC信息表						
变量	时间(年)	天数	曲线下面积(AUC)	置信区间	累计生存率	累计发病率
ERBB2	1	365	0.57633	0.3596 - 0.793	0.88793	0.11207

· 通过timeROC包计算得到每个时间点对应的曲线下面积和累计生存率
· 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算对应的曲线下面积

- 多个预测分子进行分析时可以计算其组间检验：

ROC检验表				
变量1	变量2	时间(年)	天数	p值
ERBB2	FAM241B	1	365	0.3175
ERBB2	TMEM37	1	365	0.7585
FAM241B	TMEM37	1	365	0.1103

· timeROC包提供了检验同一个时间点下不同变量的AUC之间是否有差异的方法(compare)

线

线

颜色

样式

实线

粗细

0.75pt

不透明度

1

- 颜色：线条颜色，对应 ROC 曲线的颜色，有多少个预测变量(单个预测时间)/预测时间(单个预测变量)取多少个颜色
- 线条类型：ROC 曲线的线条类型，默认为实线，可以选择虚线类型
- 线条粗细：可以选择修改 ROC 曲线线条粗细，默认为 0.75pt
- 不透明度：可以修改 ROC 曲线不透明度，默认为 1，表示完全不透明

点

点

展示

填充色

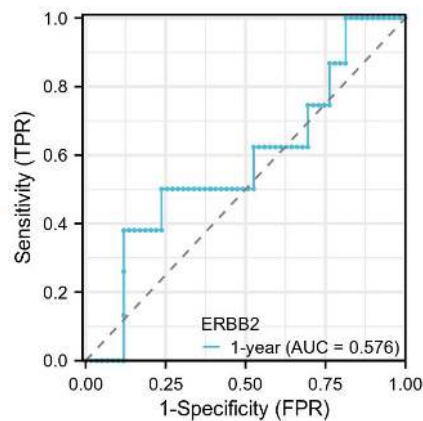
描边色

样式
圆形

大小
0.3

不透明度
1

- 展示：是否展示 ROC 曲线上的点，默认不展示，还可以选择展示，如下

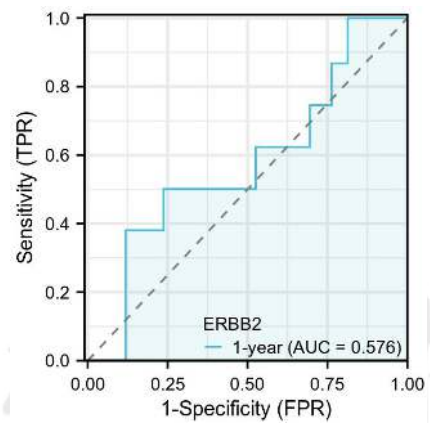


- 填充色：在选择展示点时，可以修改 ROC 曲线上点的填充色
- 描边色：在选择展示点时，可以修改 ROC 曲线上点的描边色
- 样式：在选择展示点时，可以修改 ROC 曲线上点的样式，默认为圆形，还可以同时选择正方形、菱形、三角形、倒三角形
- 大小：在选择展示点时，可以修改 ROC 曲线上点的大小
- 不透明度：在选择展示点时，可以修改 ROC 曲线上点的不透明度，默认为 1，表示完全不透明

曲线下面积



- 是否展示：可以选择是否展示曲线下面积（填充曲线下面积的部分），默认不展示，还可以选择展示，如下：



- 不透明度：可以修改曲线下面积的不透明度，前提是在展示曲线下面积

标题文本

标题	
大标题	大标题内容
x轴标题	x轴标题内容
y轴标题	y轴标题内容

- 大标题：大标题文本
- x 轴标题：x 轴标题文本
- y 轴标题：y 轴标题文本

补充：在要换行的中间插入\n。如果需要上标，可以用两个英文输入法下的大括号括住，比如 {{2}}；如果需要下标，可以用两个英文输入法下的中括号括住，比如 [[2]]

图注

图注

是否展示

☒

图注标题

图注标题内容

图注位置

默认

- 是否展示：可以选择是否展示图注信息，默认展示
- 图注标题：可以修改图注标题的内容
- 图注位置：默认为图片的右下，还可以选择右



风格



- 边框：是否添加主图的外框
- 网格：是否添加网格
- 文字大小：控制整体文字大小，默认为 7pt



图片

图片

▼

宽度 (cm)

5

高度 (cm)

5

字体

Arial

▼

- 宽度：图片横向长度，单位为 cm
- 高度：图片纵向长度，单位为 cm
- 字体：可以选择图片中文字的字体

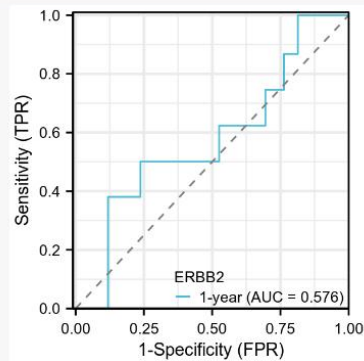


结果说明

主要结果：单个预测变量单个预测时间点

时间依赖ROC-分子

时间依赖ROC-云: ROC曲线是反映敏感性和特异性连续变量的综合指标; 时间依赖ROC则表示当预后结局随时间不断被观察时, 它的预后性能取决于预后时间t



时间依赖ROC.pdf

时间依赖ROC.tif

时间依赖ROC.pptx

分析数据.xlsx

· 横坐标表示 假阳性率(FPR)(1-特异度); 纵坐标表示 真阳性率(TPR)(灵敏度)

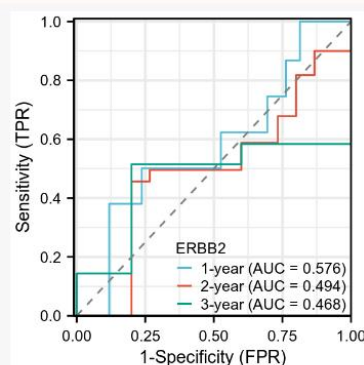
· 当一个变量(危险因素)的值是促进事件发生的趋势时, 该分子的AUC会>0.5, 此时面积越大(AUC值越接近于1)说明预后效能越好; AUC在0.5~0.7时, 预后效能较低; AUC在0.7~0.9时, 有一定预后效能; AUC在0.9以上时, 预后效能较好

· 当一个变量(保护因素)的值与事件发生的趋势相反时, 则这个分子的AUC会<0.5, 此时面积越小(AUC值越接近于0)说明预后效能越好

主要结果：单个预测变量多个预测时间点

时间依赖ROC-分子

时间依赖ROC-云: ROC曲线是反映敏感性和特异性连续变量的综合指标; 时间依赖ROC则表示当预后结局随时间不断被观察时, 它的预后性能取决于预后时间t



时间依赖ROC.pdf

时间依赖ROC.tif

时间依赖ROC.pptx

分析数据.xlsx

· 横坐标表示 假阳性率(FPR)(1-特异度); 纵坐标表示 真阳性率(TPR)(灵敏度)

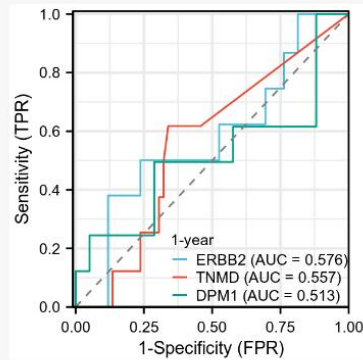
· 当一个变量(危险因素)的值是促进事件发生的趋势时, 该分子的AUC会>0.5, 此时面积越大(AUC值越接近于1)说明预后效能越好; AUC在0.5~0.7时, 预后效能较低; AUC在0.7~0.9时, 有一定预后效能; AUC在0.9以上时, 预后效能较好

· 当一个变量(保护因素)的值与事件发生的趋势相反时, 则这个分子的AUC会<0.5, 此时面积越小(AUC值越接近于0)说明预后效能越好

主要结果：多个预测变量多个预测时间点

时间依赖ROC-分子

时间依赖ROC-云: ROC曲线是反映敏感性和特异性连续变量的综合指标; 时间依赖ROC则表示当预后结局随时间不断被观察时, 它的预后性能取决于预后时间t



时间依赖ROC.pdf

时间依赖ROC.tiff

时间依赖ROC.pptx

分析数据.xlsx

· 横坐标表示 假阳性率(FPR)(1-特异度); 纵坐标表示 真阳性率(TPR)(灵敏度)

· 当一个变量(危险因素)的值是促进事件发生的趋势时, 该分子的AUC会>0.5, 此时面积越大(AUC值越接近于1)说明预后效能越好; AUC在0.5~0.7时, 预后效能较低; AUC在0.7~0.9时, 有一定预后效能; AUC在0.9以上时, 预后效能较好

· 当一个变量(保护因素)的值与事件发生的趋势相反时, 则这个分子的AUC会<0.5, 此时面积越小(AUC值越接近于0)说明预后效能越好

默认选择设置时间点中最小的时间进行分析

特殊参数

分子

ERBB2[ENSG00000141736.14] ×

TNMD[ENSG00000000005.6] ×

DPM1[ENSG000000000419.13] ×

预测时间

时间1 1

时间2 2

时间3 3

时间单位 年

补充结果：单个预测变量多个预测时间点的情况

统计描述

变量	时间(年)	天数	事件数(截至当前时间)	生存数(截至当前时间)	删失数(截至当前时间)
ERBB2	1	365	8	59	15
ERBB2	2	730	15	15	52
ERBB2	3	1095	17	5	60

- 如果某个时间点的事件数为0，则该时间点无法进行分析和可视化
- 如果某些时间点的事件数、生存数、删失数相同，建议增加时间间隔（如果时间点大于数据的最大时间，这些结果也会相同）
- 注意：数据默认在分析前统一删除变量缺失的样本

这里的表格提供不同时间点的事件情况

- 如果某个时间点的事件数为 0，则该时间点无法进行分析和可视化
- 如果某些时间点的事件数、生存数、删失数相同，建议增加时间间隔（如果时间点大于数据的最大时间，这些结果也会相同）
- 数据默认在分析前统一删除变量缺失的样本

ROC结果表

变量	时间(年)	天数	最佳阈值(cut-off)	敏感度	特异度	阳性预测值	阴性预测值
ERBB2	1	365	5.994602	0.50112	0.76271	0.21046	0.92374
ERBB2	2	730	6.407936	0	0.8	0	0.59953
ERBB2	3	1095	5.581761	0.58381	0	0.33684	0

ROC结果表.xlsx

- 先通过（敏感度+(1-特异度)）得到最大对应的第一个cut_off作为最佳阈值（cut_off），再通过timeROC包计算得到最佳阈值（cut_off）下的部分ROC相关信息和数据
- timeROC包值提供单个cut_off值的灵敏度、特异度、阳性预测值和阴性预测值的结果，因此无法提供其他ROC相关的指标结果
- 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算则无法进行相关的时间依赖ROC分析

这里的表格为 ROC 结果表格，提供不同时间点的 ROC 曲线下面积的结果及其其他指标

- 先通过（敏感度+(1-特异度)）得到最大对应的第一个 cut_off 作为最佳阈值（cut_off），再通过 timeROC 包计算得到最佳阈值（cut_off）下的部分 ROC 相关信息和数据
- timeROC 包值提供单个 cut_off 值的灵敏度、特异度、阳性预测值和阴性预测值的结果，因此无法提供其他 ROC 相关的指标结果

- 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算对应的曲线下面积

AUC信息表

变量	时间(年)	天数	曲线下面积(AUC)	置信区间	累计生存率	累计发病率
ERBB2	1	365	0.57633	0.3596 - 0.793	0.88793	0.11207
ERBB2	2	730	0.4937	0.2636 - 0.7238	0.65173	0.34827
ERBB2	3	1095	0.46807	0.1899 - 0.7463	0.53475	0.46525

· 通过timeROC包计算得到每个时间点对应的曲线下面积和累计生存率

· 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算对应的曲线下面积

这里的表格为 AUC 信息表，提供不同预测时间点的信息

- 通过 timeROC 包计算得到每个时间点对应的曲线下面积和累计生存率
- 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算对应的曲线下面积



补充结果：多个预测变量多个预测时间点的情况（并且计算置信区间的前提下）

统计描述

变量	时间(年)	天数	事件数(截至当前时间)	生存数(截至当前时间)	删失数(截至当前时间)
ERBB2	1	365	8	59	15
TNMD	1	365	8	59	15
DPM1	1	365	8	59	15

- 如果某个时间点的事件数为0，则该时间点无法进行分析和可视化
- 如果某些时间点的事件数、生存数、删失数相同，建议增加时间间隔（如果时间点大于数据的最大时间，这些结果也会相同）
- 注意：数据默认在分析前统一删除变量缺失的样本

ROC结果表

变量	时间(年)	天数	最佳阈值(cut-off)	敏感度	特异度	阳性预测值	阴性预测值
ERBB2	1	365	5.99460235	0.50112	0.76271	0.21046	0.92374
TNMD	1	365	0.04222407	0.61776	0.66102	0.187	0.93198
DPM1	1	365	7.10786150	0.49475	0.71186	0.17812	0.91778

ROC结果表.xlsx

- 先通过（敏感度+(1-特异度)）得到最大对应的第一个cut_off作为最佳阈值（cut_off），再通过timeROC包计算得到最佳阈值（cut_off）下的部分ROC相关信息和数据
- timeROC包提供单个cut_off值的灵敏度、特异度、阳性预测值和阴性预测值的结果，因此无法提供其他ROC相关的指标结果
- 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算则无法进行相关的时间依赖ROC分析

AUC信息表

变量	时间(年)	天数	曲线下面积(AUC)	累计生存率	累计发病率
ERBB2	1	365	0.57633	0.88793	0.11207
TNMD	1	365	0.55747	0.88793	0.11207
DPM1	1	365	0.51317	0.88793	0.11207

- 通过timeROC包计算得到每个时间点对应的曲线下面积和累计生存率
- 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算对应的曲线下面积

计算置信区间：

ROC检验表

变量1	变量2	时间(年)	天数	p值
ERBB2	TNMD	1	365	0.9081
ERBB2	DPM1	1	365	0.7450
TNMD	DPM1	1	365	0.7851

- timeROC包提供了检验同一个时间点下不同变量的AUC之间是否有差异的方法(compare)

- timeROC 包提供了检验同一个时间点下不同变量的 AUC 之间是否有差异的方法 (compare)
- 当预测变量相同时，其对应的 p 值没有意义，即为 NA



方法学

统计分析和可视化均在 R 4.2.1 版本中进行

涉及的 R 包：timeROC（用于分析），ggplot2 包（用于可视化）

处理过程：

- (1) 通过 timeROC 包对云端数据进行分析
- (2) 使用 ggplot2 包将所得分析结果进行可视化



如何引用

生信工具分析和可视化用的是 R 语言，可以直接写自己用 R 来进行分析和可视化即可，可以无需引用仙桃，如果想要引用仙桃，可以在致谢部分 (Acknowledge) 致谢仙桃学术 (www.xiantao love)。

方法学部分可以参考对应说明文本中的内容以及一些文献中的描述。

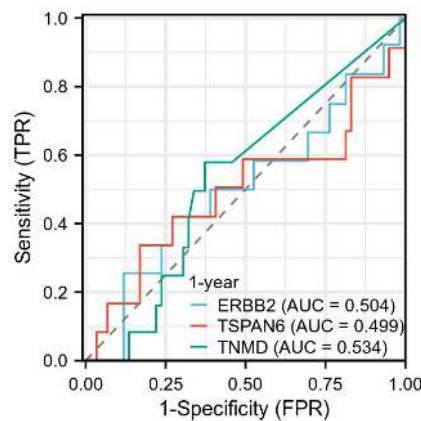


常见问题

1. 多个预测变量的时候是什么样子的，多个预测变量的时候还能预测多个时间节点吗？

答：

当选择了多个分子时，只会分析当前预测时间中对应设置的最小的时间



2. 为什么有多个曲线，没有给出统计检验的 p 值？

答：同一个预测变量不同时间点没办法进行统计检验（timeROC 包不提供）；只有不同预测变量同一个时间点才会有统计检验的结果

3. 为什么 AUC 会小于 0.5

答：timeROC 包固定了（无法修改）在一个时间段内发生事件的作为病例组，所以当有一个分组数值的趋势与预后结局趋势相反（类似保护因素 $HR < 1$ ）时，则会出现 $AUC < 0.5$ ，此时 AUC 越接近 0，说明预测效能越好