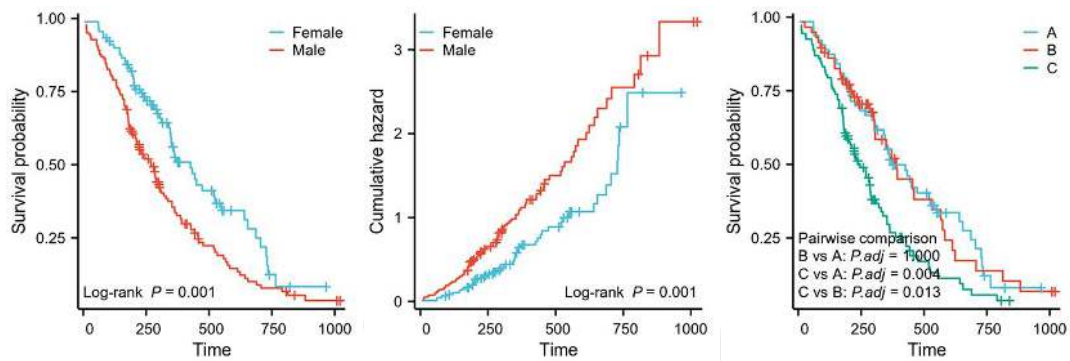


临床意义 - 生存曲线



网址: <https://www.xiantao.love>



更新时间: 2023.03.29

目录

基本概念	3
应用场景	4
结果解读	5
数据格式	6
参数说明	7
分组	7
统计	8
类型	10
线	11
删失数据	11
置信区间	12
风险表格	12
标题文本	13
图注(Legend)	13
坐标轴	14
风格	15
图片	15
结果说明	16
主要结果	16
补充结果	16
方法学	18
如何引用	19
常见问题	20

基本概念

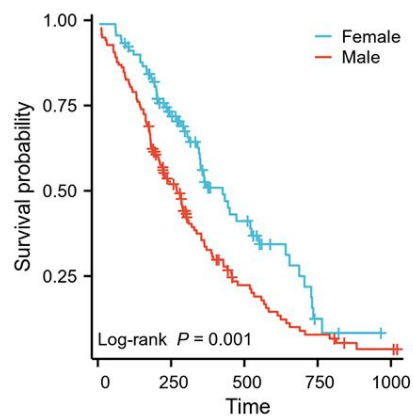
- 生存曲线（也称 Kaplan-Meier 曲线）：可以描述各组患者的生存状况或者各组实验动物的存活情况
- 生存时间：从规定的起始事件开始到失效事件出现所持续的时间。对于失访者，是失访前最后一次随访的时间。
- 终点事件/终点结局：医学研究中可以是患者死亡，也可以是疾病的发生、某种治疗的反应、疾病的复发等。与之对应的起始事件可以是疾病的确诊、某种治疗的开始等。
- 删失/截尾（Censoring）：由于某些原因在随访中并没有观测到终点结局而不知道确切的生存时间，此部分数据即删失数据。常见原因有失访、患者退出试验、事件发生是由于非研究性疾病（如研究病人发生脑卒中后的生存时间，结果病人因为车祸死亡）、研究结束时研究对象仍未发生失效事件。删失数据的生存时间为起始事件到截尾点所经历的时间。
- 中位生存时间（Median Survival Time）：中位生存时间又称半数生存期，表示恰好一半个体未发生终点事件的时间，生存曲线上纵轴 50% 对应的时间。如果删失或者截尾数据较多，预后较好，则可能无法计算得到对应的中位生存时间。
- 生存分析的方法：
 - 非参数法：寿命表、Kaplan-Meier(乘积极限法 Product limit method/检验方法：对数秩 (Log rank)、Breslow、Tarone-Ware)等
 - 半参数法：Cox 回归（需要满足 Cox 比例风险假设）
 - 参数法

- 备注：Log rank 方法 和 Cox 方法都有很广泛的应用，一般两者选其一即可。
- PH 假设：比例风险（Proportional hazards）假定。Cox 模型应用的前提条件。基本假设为：协变量对生存率的影响不随时间的改变而改变，即风险比值 $h(t)/h_0(t)$ 为固定值。而在实际进行生存分析的过程中，有些自变量对风险函数（事件发生概率）的影响会随时间的变化而变化，因此在构建 Cox 回归模型之前，必须对 PH 假定进行判定，只有 PH 假定得到满足时，Cox 回归模型的结果才有意义。
- 风险比（Hazard Ratio, HR）：两个风险率的比值。当 $HR > 1$ 时，说明相对于对照组，实验组（研究对象）是一个危险因素；当 $HR < 1$ 时，说明相对于对照组，实验组（研究对象）是一个保护因素；当 $HR = 1$ 时，说明研究对象对生存时间不起作用。
- 统计类型：（可根据需要选择，一般选择展示累积生存）
 - 累积生存：时间段内生存概率的累积。
 - 累积风险：时间段内风险概率的累积
 - 累积事件：时间段内事件发生率。

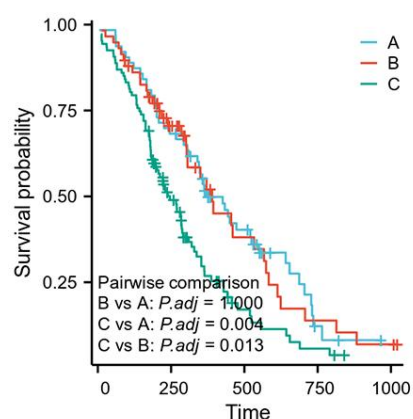
应用场景

生存曲线主要用于描述 受试或者研究对象 在一段时间内发生事件（存活 or 死亡 / 是否复发等）的情况。

结果解读



- 生存曲线的横坐标是观察时间，纵坐标一般是生存率。曲线上的每一个点代表了在该时间点上病人的生存率。
- 在坐标轴 (0, 1) 点上，由于才开始随访 (X 轴为 0)，此时没有患者死亡，所以两组患者的生存率都是 1 (100%)。
- 代表女性的蓝色曲线始终在代表男性的红色曲线上方，表明在同一时间节点上，女性存活率较男性存活率高，女性这个因素预后更佳。
- 曲线上的“+”用于标识删失数据。



- 多组数据（最多支持 4 组）可提供分组比较的结果。图中的统计结果说明，相对于 A 和 B，C 组有显著的统计学差异（校正后 p 值 < 0.05 ），结合图中的趋势来看，C 组的预后相对较差。

数据格式

1	status	time	sex
2	1	310	Female
3	1	361	Female
4	1	654	Female
5	1	728	Female
6	1	61	Female
7	1	81	Female
8	1	520	Female
9	1	473	Female
10	1	107	Female
11	1	122	Female
12	0	965	Female
13	1	731	Female
14	1	459	Female

1	event	time	expr
2	2	492	1.6906665
3	2	394	1.7057732
4	2	2681	1.8791092
5	1	204	1.9987961
6	2	138	2.1211438
7	2	585	2.1327881
8	2	771	2.2738583
9	1	908	2.3279909
10	1	705	2.3441711
11	2	1481	2.3484183
12	1	1337	2.389901
13	1	1012	2.3905073

数据至少 2 列，最多 3 列：

- 第一列是**事件发生情况**，用 0 和 1 表示，0 表示未发生事件（删失），1 表示发生了事件。例如，事件可以定义为 死亡，当受试发生了死亡，该受试的事件就定义为 1，当受试未发生死亡（删失），该受试的事件就定义为 0。也可以用 1, 2 表示，此时 1 代表未发生事件（删失），2 代表是发生了事件。
- 第二列是**具体时间**，可以是以 天、月、年 为单位。
- 第三列（可提供）为变量，可以是**二分类或者多分类变量**，也可以是**纯数值类型的变量（同一个数字不能超过 30%）**。如果不提供，则只能绘制整体（一组）的结果。
 - **分类变量** 默认是排在前面的分组作为参考，最大支持 4 分类。在“分组”选项卡中，可以选择其他分组作为参考组。
 - **数值类型的变量** 可以在右侧参数中调整分组的参数从而实现转换成二分类分组。



参数说明

(说明：标注了颜色的为常用参数。)

分组



➤ **分组**：可以选择分组的信息

- 如果上传数据的第三列是**二分类或者是多分类**，那么分组选项这里会出现左侧图的选项，可以选择参考组（影响 HR 值的计算）。整体（所有数据 1 个组）即把整个预后数据当做 1 个组，绘制一条生存曲线、
- 如果上传数据的第三列是**数值类型**，那么分组选项这里会出现右侧图的选项，其中 0-50 vs 50-100 是中位数分组。0-33 vs 66-100 是以数值按从小到大进行排序，从前往后的**取前 33%作为低**，**取后 33%为高**。其余分组以此类推。p 值最小分组为 survminer 包提供的一种方法，即将数值变量的不同数值作为 cut-off，每个数值均尝试作为分组的 cut-off 值，将每次尝试得到的 p 值进行排序，得到的最小的 p 值对应的 cut-off 值作为分组的 cut-off。这个类似于 KMplot 网站中的 best-separate 分割。（建议是先尝试中位数分组，其次是尝试 p 值最小。）

统计

统计

统计方法 Logrank检验

标注位置 左下

标注颜色

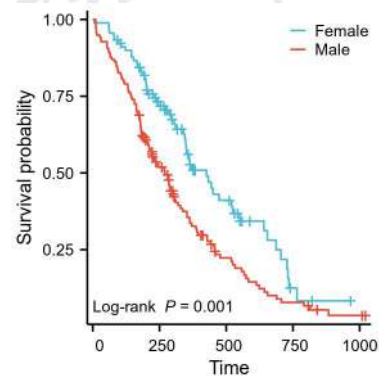
- **统计方法**: 在生存分析中 Kaplan-Meier(K-M)法可以估计生存概率, Log-rank 检验可以比较两条或多条生存曲线, 用 Cox 比例风险回归模型可以分析结局转归的影响因素。Cox 比例风险回归模型用于调查患者和一个或多个预测变量的存活时间之间的关联回归模型, 并且也能够同时评估多个因素对生存的影响。(两种方法均有很广的使用, Cox 需要满足 ph 假设), **当选择 Cox 回归分析时, 会在图中进行 HR 值标注**, 如下:

统计

统计方法 Logrank检验

标注位置 左下

标注颜色

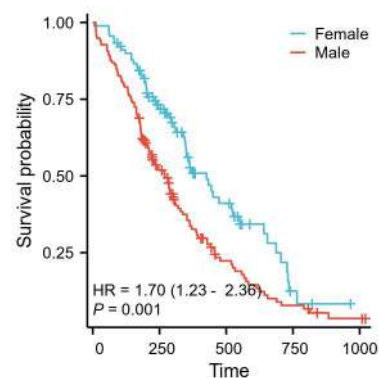


统计

统计方法 Cox回归

标注位置 左下

标注颜色



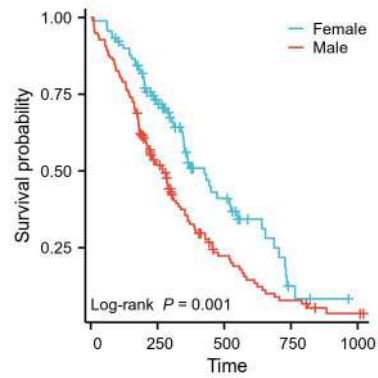
- **标注位置**：可以选择统计学标注的文字显示在图片的位置，如下：

统计

统计方法 Logrank检验

标注位置 左下

标注颜色

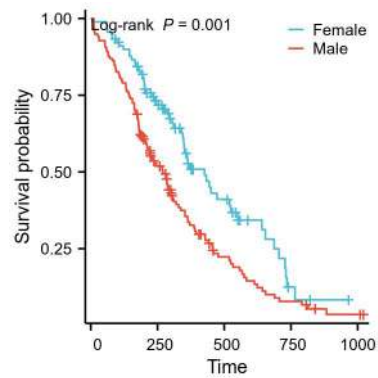


统计

统计方法 Logrank检验

标注位置 左上

标注颜色



- **标注颜色**：可以修改统计学标注文字的颜色

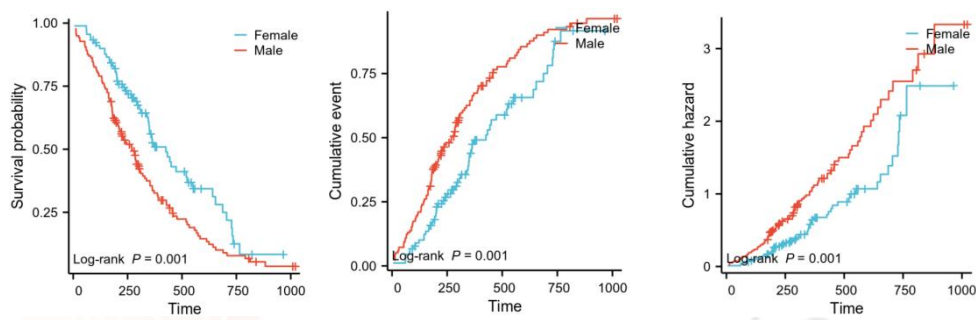
类型

类型

类型

生存概率(Su

- 类型：可以选择生存概率、累积事件、累积风险三种类型。最常用的为生存概率。



线



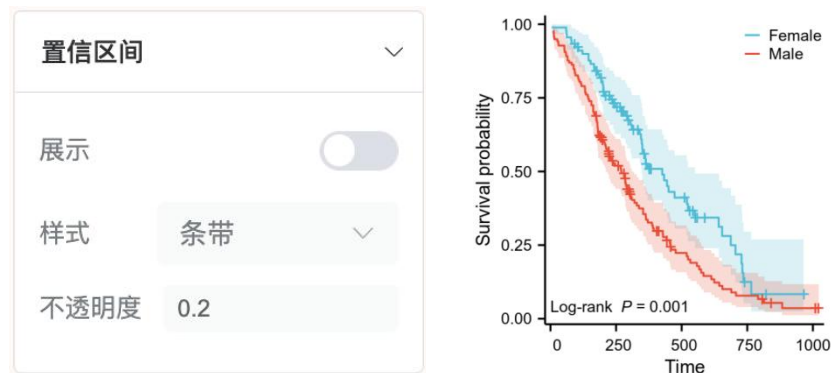
- **颜色**: 生存曲线对应的颜色，有几条线就会取几个颜色。这里的颜色受全局参数影响。
- **样式**: 可以选择实线或者虚线。
- **粗细**: 线的粗细

删失数据



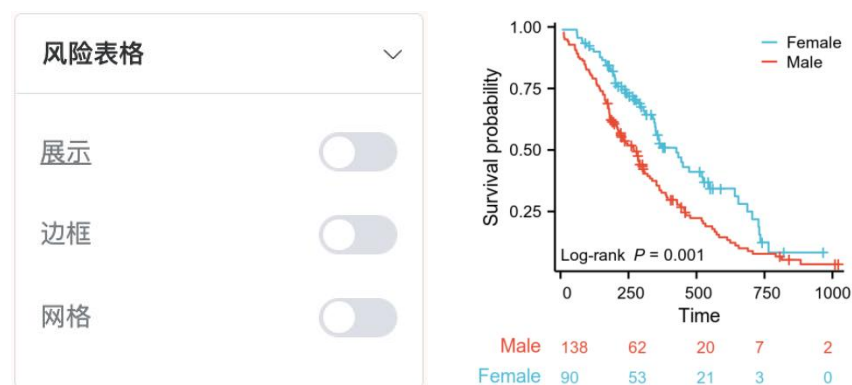
- **展示**: 是否展示删失数据，图中标注了 的就是删失数据，一般默认展示
- **颜色**: 删失标注的颜色，默认是和线同色，可以根据需要进行修改
- **大小**: 的大小

置信区间



- 展示：是否展示每个分组每个时间点累积生存率的置信区间
- 样式：可选条带和虚线
- 透明度：设置置信区间的透明度，0 为完全透明，1 为完全不透明。

风险表格



风险表格记录了各个时间点上还在随访的人数

- 展示：是否展示风险表格，默认是不展示。开启后如右图下部分所示
- 外框：风险表格部分是否有外框，只有开启展示才有作用
- 网格：风险表格部分是否有网格，只有开启展示才有作用

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(Legend)

图注 ▼

是否展示

☒

图注标题

图注标题内容

图注位置

默认 ▼

- 展示：是否展示图注
- 图注位置：可选（图中）右上、右下、左上、左下，（图外）右、上，默认为右上。

- 图注标题：可以添加图注标题

坐标轴

坐标轴
▼

x轴范围+刻度 逗号隔开

y轴范围+刻度 逗号隔开

- X 轴范围+刻度：（注意：x 轴对应的范围必须是要在数据的随访时间内，可以根据需要修改成想要的时间刻度或者是修改范围）
 - 如果只是想要修改范围，可以只输入两个范围值，比如 0, 500。如果是想要同时修改范围和刻度，可以输入范围+刻度，比如 0, 100, 200, 300, 400, 500, 500.
- Y 轴范围+刻度：（注意：范围的修改不可以过大或者过小）
 - 如果只是想要修改范围，可以只输入两个范围值，比如 0.2, 1。如果是想要同时修改范围和刻度，可以输入范围+刻度，比如 0, 0, 0.1, 0.2, 0.5, 0.5.

风格



风格

边框 ☐

网格 ☐

文字大小 7pt

- 外框：是否添加主图外框
- 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

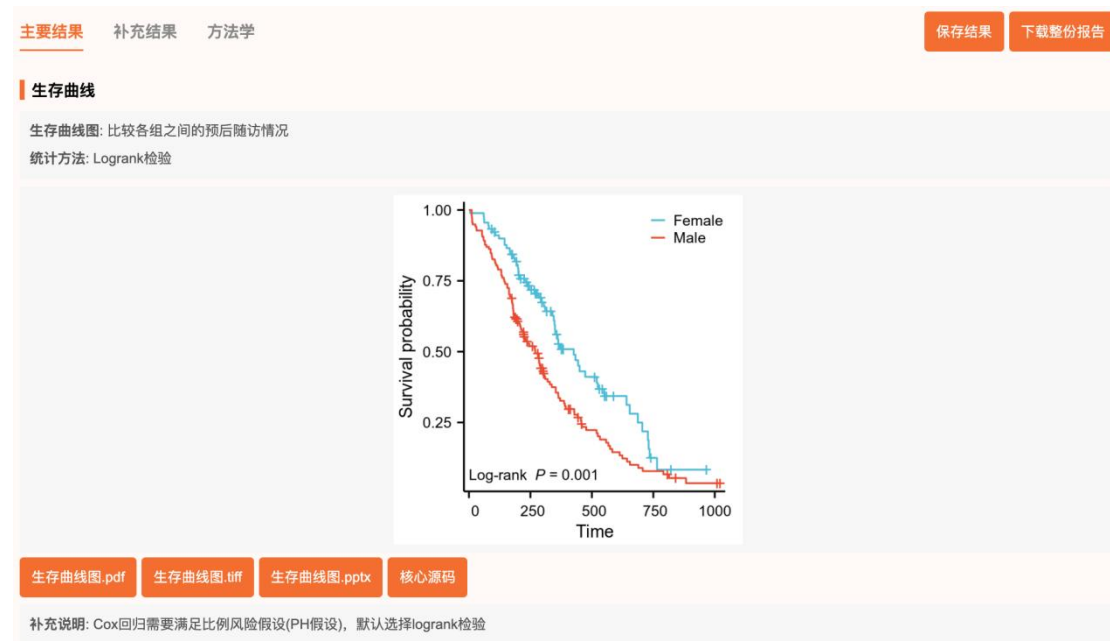
高度 (cm) 5

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



补充结果

统计描述

分组	数目	总事件数	总删失数	总删失比例	中位生存时间	中位生存时间置信区间
Female	90	53	37	41.1%	426	348-550
Male	138	112	26	18.8%	270	212-310

累积生存率.xlsx

备注: 中位生存时间的置信区间如果有?, 则代表 分组中样本较少 或者是 随访时间不足 或者是 预后相对较好无法计算出来对应的上限或者下限

这里提供统计描述的表格, 包含删失情况和中位生存时间统计情况(注意: 如果数据预后情况较好, 则对应的中位生存时间可能计算不出来, 包括中位生存时间置信区间)

这里提供每个分组每个时间点的累积生存率情况表格xlsx 下载（一个 sheet 代表一个分组情况）：（可以查看某个时间节点的累积生存率，比如 1 年、5 年）

	A	B	C	D	E	F	G
1	time	n.risk	n.event	survival	std.err	lower 95%	upper 95%
2	5	90	1	0.9888889	0.0110492	0.9674682	1
3	60	89	1	0.9777778	0.0155379	0.9477934	1
4	61	88	1	0.9666667	0.0189215	0.9302835	1
5	62	87	1	0.9555556	0.0217228	0.9139143	0.9990942
6	79	86	1	0.9444444	0.0241452	0.8982868	0.9929738
7	81	85	1	0.9333333	0.0262937	0.8831956	0.9863173
8	95	83	1	0.9220884	0.0282794	0.8682947	0.9792147
9	107	81	1	0.9107045	0.0301346	0.8535164	0.9717245
10	122	80	1	0.8993207	0.0318356	0.8390396	0.9639328
11	145	79	2	0.8765531	0.0348634	0.8108176	0.947618
12	153	77	1	0.8651693	0.0362215	0.7970111	0.9391563
13	166	76	1	0.8537855	0.0374911	0.7833775	0.9305216
14	167	75	1	0.8424017	0.038681	0.7698998	0.9217311
15	182	71	1	0.8305369	0.0399144	0.7558775	0.9125705
16	186	70	1	0.8186721	0.0410698	0.7420075	0.9032576
17	194	68	1	0.8066328	0.0421935	0.7280329	0.8937184
18	199	67	1	0.7945935	0.0432473	0.7141948	0.8840428
19	201	66	2	0.7705149	0.0451641	0.6868904	0.8643202
20	208	62	1	0.7580872	0.0461128	0.6728857	0.8540771

- time: 上传数据的每个时间
- n.risk: 每个时间点对应的人数
- n.event: 每个时间点上发生事件的人数
- survival: 当前时间点的累积生存率。
- std.err: 当前时间点的累积生存率的标准误
- lower 95%: 当前时间点的累积生存率的 95%置信区间下限
- upper 95%: 当前时间点的累积生存率的 95%置信区间上限

比例风险假设(PH)

Cox回归应用的前提是要求自变量满足等比例风险假设($P > 0.05$)，即自变量的风险不会随着时间改变而改变，若不满足，则不适合用Cox回归进行检验

Logrank检验没有要求满足比例风险假设，当不满足比例风险假设时可以临时选用Logrank检验，但是最严谨的是采用RMST(Restricted mean survival time)方法，当前模块无法兼容采用RMST方法

统计量(卡方值)	自由度	p值
2.8632	1	0.0906

如果p值小于0.05，则说明变量不满足比例风险假设，此时可以选用Logrank检验

这里提供风险比例假设检验情况，可以根据这个判断是否选择 cox 回归方法

统计分析

同时提供Log-rank和Cox回归的检验结果，Cox回归应用需要满足风险比例假设(PH假设)

方法	统计量	HR	置信区间	p值
Log-rank	10.33	1.69	1.25-2.30	0.0013
Cox回归	10.63	1.70	1.23-2.36	0.0015

参考组(Reference): Female

此表为利用两种统计学方法进行统计分析的结果，可以按需提取需要的结果。

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：survminer 包（用于可视化），survival 包（用于生存资料的统计分析）

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

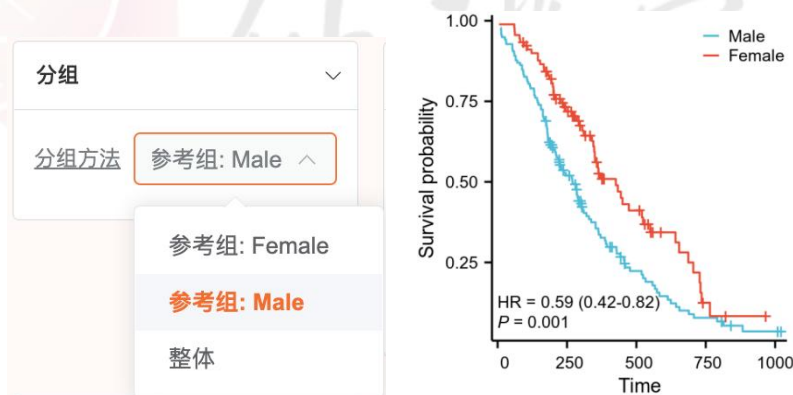
1. 如何 更改 Cox 分析结果中的 reference 组 (HR 值相反) ?

答:

工具会以分组中首行出现的组默认作为 reference 组:



如果选择另外一个组作为参考组:



2. 应该选择 log rank 方法 还是 Cox 方法?

答:

两种方法均可, Cox 回归需要满足 PH 假设, 当不满足 PH 假设时, 建议是选择 Log rank 方法。

3. 如何查看某个分组 1 年生存率（或者某年生存率）？

答：

可以在补充结果下载“累积生存率.xlsx”表格：

统计描述						
分组	数目	总事件数	总删失数	总删失比例	中位生存时间	中位生存时间置信区间
Male	138	112	26	18.8%	270	212-310
Female	90	53	37	41.1%	426	348-550

累积生存率.xlsx

备注：中位生存时间的置信区间如果有？，则代表 分组中样本较少 或者是 随访时间不足 或者是 预后相对较好无法计算出来对应的上限或者下限

	A	B	C	D	E	F	G	H
1	time	n.risk	n.event	survival	std.err	lower 95%	upper 95% CI	
2	5	90	1	0.9888889	0.0110492	0.9674682	1	
3	60	89	1	0.9777778	0.0155379	0.9477934	1	
4	61	88	1	0.9666667	0.0189215	0.9302835	1	
5	62	87	1	0.9555556	0.0217228	0.9139143	0.9990942	
6	79	86	1	0.9444444	0.0241452	0.8982868	0.9929738	
7	81	85	1	0.9333333	0.0262937	0.8831956	0.9863173	
8	95	83	1	0.9220884	0.0282794	0.8682947	0.9792147	
9	107	81	1	0.9107045	0.0301346	0.8535164	0.9717245	
10	122	80	1	0.8993207	0.0318356	0.8390396	0.9639328	
11	145	79	2	0.8765531	0.0348634	0.8108176	0.947618	
12	153	77	1	0.8651693	0.0362215	0.7970111	0.9391563	
13	166	76	1	0.8537855	0.0374911	0.7833775	0.9305216	
14	167	75	1	0.8424017	0.038681	0.7698998	0.9217311	
15	182	71	1	0.8305369	0.0399144	0.7558775	0.9125705	
16	186	70	1	0.8186721	0.0410698	0.7420075	0.9032576	
17	194	68	1	0.8066328	0.0421935	0.7280329	0.8937184	
18	199	67	1	0.7945935	0.0432473	0.7141948	0.8840428	
19	201	66	2	0.7705149	0.0451641	0.6968904	0.8643202	
20	208	62	1	0.7580872	0.0461138	0.6728857	0.8540771	

这里面一个 sheet 代表一个分组的情况，time 的单位与上次数据的时间的单位一致，这里示例数据的单位为天，所以在表格中找 1*365 附近的值，可以从表格中找到时间为 363 对应的内容：

	A	B	C	D	E	F	G
1	time	n.risk	n.event	survival	std.err	lower 95%	upper 95% CI
34	361	33	1	0.5404437	0.0592000	0.408905	0.6727944
35	363	32	1	0.526463	0.0597369	0.4214863	0.6575855
36	371	30	1	0.5089143	0.0602682	0.4034977	0.6418716

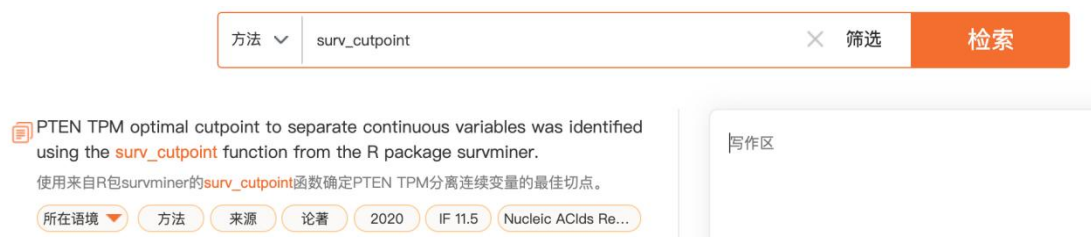
此时的生存率为 52.6%，这个就是 Female 组 1 年生存率的情况。

4. 如何理解 p 值最小分组，如何在方法学中进行说明？

答：

p 值最小分组为 survminer 包提供的一种方法，即将数值变量的不同数值作为 cut-off，每个数值均尝试作为分组的 cut-off 值，将每次尝试得到的 p 值进行排序，得到的最小的 p 值对应的 cut-off 值作为分组的 cut-off。这个类似于 KMplot 网站中的 best-separate 分割。

方法学中在写作工具中限定“方法学”搜索“surv_cutpoint”：



或者以下这些例子（请勿直接复制）：

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7841432/>
 - The “surv_cutpoint” function in the survminer R package was performed to search the best split by verifying all potential cut points.
- <https://www.frontiersin.org/articles/10.3389/fmolb.2021.608369/full>
 - The “surv_cutpoint” function in the survminer package was used to determine the optimal cut-off value of the risk score.

5. 如何绘制一组生存曲线？

答：



当上传的数据只有两列（1 列是事件、1 列是时间），则只能绘制 1 组生存曲线。

当上传的数据有三列（1 列事件、1 列是时间、1 列是分组），则可以在主要参数“分组”选项卡中选择“整体”。

