

R语言实例解析肿瘤治疗药物经济学评价的分区生存模型



刘家一¹, 李 伟²

1. 中国人民大学公共管理学院 (北京 100872)

2. 北京医院麻醉科 (北京 100005)

【摘要】 分区生存模型 (partitioned survival model, PSM) 作为药物经济学 (pharmacoeconomics, PE) 评价的一种决策模型, 因其简单直观的特点, 以及其分析数据直接源于临床文献且避免了复杂的状态间转移概率计算, 在国外肿瘤治疗 PE 评价中得到广泛应用。R 语言在进行肿瘤治疗药物 PE 评价的 PSM 分析中, 以其快捷、直观且易于复现结果的特点, 已在欧美研究者的多年应用经验中形成了多种功能专用程序包, 极大地提高了研究效率。然而, 我国采用 R 语言进行肿瘤治疗药物的 PSM 分析相关研究较少。本研究结合实际案例详细展示基于 R 语言进行肿瘤治疗药物 PSM 分析的完整操作流程, 以期为该领域的研究者提供参考。

【关键词】 R 语言; 药物经济学评价; 卫生经济学评价; 分区生存模型; 肿瘤治疗药物

Partitioned survival model of pharmacoeconomics evaluation of oncology therapy in R language

LIU Jiayi¹, LI Wei²

1. School of Public Administration and Policy, Renmin University of China, Beijing 100872, China

2. Department of Anesthesiology, Beijing Hospital, Beijing 100005, China

Corresponding author: LI Wei, Email: 13810004011@163.com

【Abstract】 The partitioned survival model (PSM), utilized in pharmacoeconomics (PE) evaluation for decision-making models, is widely applied in PE evaluation of international oncology treatments due to its simplicity, intuitiveness, and direct data analysis from clinical literature, circumventing complex inter-state transition probability calculations. The R language, known for its efficiency, intuitive nature, and reproducibility of results, has been equipped with various specialized packages through years of application by researchers in Europe and America, significantly enhancing research productivity. However, there were few studies about the use of R language for PSM analysis of oncology drug treatments in China. This article aims to present a comprehensive operational procedure for PSM analysis of oncology drug treatments using R language, combined with practical cases, and to offer a reference for researchers in the field.

【Keywords】 R language; Pharmacoeconomics evaluation; Health economic evaluation; Partitioned survival model; Oncology treating drugs

DOI: 10.12173/j.issn.1004-4337.202401034

通信作者: 李伟, Email: 13810004011@163.com

<https://slyyx.whuzhmedj.com/>

分区生存模型 (partitioned survival model, PSM) 是国外肿瘤治疗药物经济学 (pharmacoeconomics, PE) 评价中应用较多的一种决策模型, 用于肿瘤药物的报销决定^[1-2]。PSM 已在英国国家卓越健康与护理研究所 (National Institute for Health and Care Excellence, NICE) 技术评估 (Technology Appraisal, TA) 计划和加拿大药物与卫生技术局 (Canadian Agency for Drugs and Technologies in Health, CADTH) 评价中广泛使用, 是目前 NICE 评估晚期或转移性癌症干预措施的最常用方法^[1,3]。与国内研究者常用的状态转移模型相比, PSM 具有简洁、直观的特点, 它直接利用肿瘤治疗药物临床报告中的无进展生存期 (progression free survival, PFS)、总生存期 (overall survival, OS) 数据, 避免在数据有限的情况下计算并验证复杂的状态间转移概率^[4-8]。

在 PSM 应用中, 尽管 TreeAge 等软件在建模方面具有一定优势, 但其固有设置要求定义状态间转移概率, 限制了其在 PSM 中的应用^[9]。相较之下, R 语言在药物成本效用分析中灵活性更强, 专用程序包可直接应用 PSM, 同时提供了丰富的数据处理与可视化功能。国际研究者已在不断尝试和探索新的模型结构和变量选择方法过程中产生了不同功能的专用程序包, 如健康经济模型分析的 heemod^[10-11]、提供灵活生存分析方法的 flexsurv^[12]、集成了健康经济模型模拟分析的 hesim^[13]、标准生存数据分析的 survival^[14]、结合了生存分析与健康经济评估的 survHE^[15-16]、执行贝叶斯成本效益分析的 BCEA^[17] 等。利用 R 语言进行肿瘤治疗药物的 PSM 分析, 能够快速、直观地复现结果, 在缺乏患者个体数据 (individual patient data, IPD) 的情况下, 研究者可直接提取临床试验报告中公开的 PFS、OS 曲线数据^[13,18], 以重新构建 IPD 并完成整个 PE 评价, 但这要求研究者具有数理统计、健康经济分析与编程的综合知识背景, 国内采用 R 语言进行 PSM 分析的应用报道不多, 且缺乏具体实施案例。本研究以帕博利珠单抗治疗非小细胞肺癌 (non-small cell lung cancer, NSCLC) 的 KEYNOTE189 临床试验 (简称 KN189)^[19-20] 为例, 实现从数字化提取 PFS、OS 曲线数据, 然后由 R 语言实现 PE 分析与评价的全过程, 以期研究人员利用 R 语言进行 PE 评价提供参考。

1 临床试验数据提取

基于帕博利珠单抗 Pembrolizumab KN189 临床试验数据^[19-20] (图 1、图 2), 使用 GetDataW 2.26.0.20 软件分别对其 OS、PFS 曲线数据进行提取, 同时提取 number at risk 数据, 一并保存为 xlsx 文件。针对提取后数据进行整理, 使后一个时间点的生存率小于或等于前一个时间点数值。

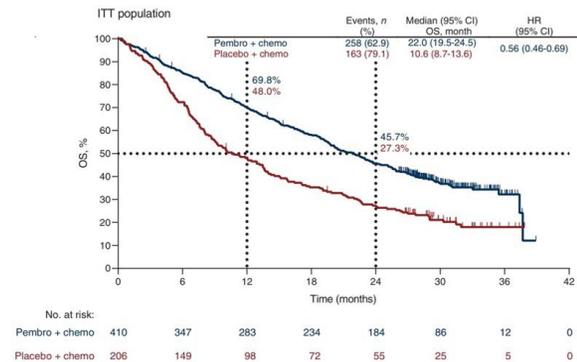


图1 KN189临床试验OS曲线数据
Figure 1. OS curve data of KN189

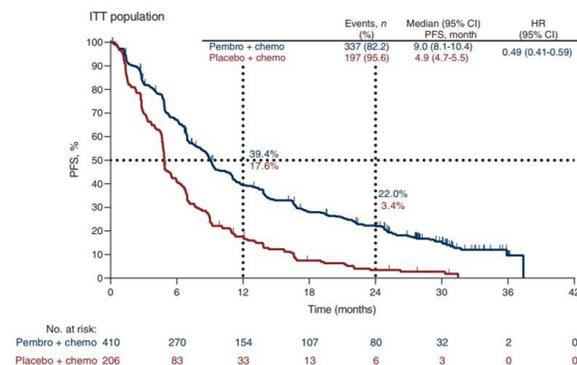


图2 KN189临床试验PFS曲线数据
Figure 2. PFS curve data of KN189

2 重构IPD

2.1 R语言环境设置

打开 Rstudio 2023.09.1+494 软件, 设定工作路径, 安装并下载软件包: survial (v 3.5-7)、survHE (v 2.0.1)、IPDfromKM (v 0.1.10)、readbitmap (v 0.1.5)、survminer (v 0.4.9)、flexsurv (v 2.2.2)、data.table (v 1.15.2)、hesim (v 0.5.4)、openxlsx (v 4.25.2)。

2.2 定义数据集

基于 KN189 试验方案，将对照组命名为 cemo，帕博利珠组定义为 comb，OS 曲线命名为 surv，PFS 曲线命名为 pfs。分别通过 read.xlsx() 函数读取第 1 步中提取的 xlsx 文件，整理并保存为如下数据集：survcomb、survcemo、nriskcomb、nriskcemo、pfscomb、ppscemo、nsirkcomb，表 1、表 2 分别展示了 survcomb 与 nriskcemo 数据集的数据结构。此处仅示例 OS 曲线的数据重构过程，PFS 曲线与 OS 曲线的数据提取过程类似。

表1 KN189试验comb组OS曲线数据
Table 1. OS curve data of the comb group in KN189

生存时间顺序	生存时间 (周)	对应生存率 (%)
1	0.0	100.0
2	0.7	100.0
3	1.3	100.0
4	1.8	99.5
5	2.4	98.8
6	3.1	98.1
.....		
286	167.0	12.4
287	167.9	12.4
288	169.0	12.4
289	169.6	12.4
290	170.3	12.4
291	171.0	11.9

表2 KN189试验cemo组OS曲线风险人数数据
Table 2. OS curve data of the cemo group in KN189

由试验周期划分的区间	KN189汇报周期 (周)	对应周期的时间风险人数
1	0.0	206
2	26.4	149
3	52.8	98
4	79.2	72
5	105.6	55
6	132.0	25
7	158.4	5
8	184.8	0

2.3 预处理与重构IPD数据

2.3.1 预处理IPD对象格式

使用 preprocess() 函数预处理为 IPD 对象适当格式，R 语言代码如下：

```
pre_comb<- preprocess (dat=survcomb,
trisk=nriskcomb$time, nrisk=nriskcomb$nrisk,
totalpts=410)
```

totalpts 为指定试验的初始人数；

```
pre_cemo<- preprocess (dat=survcemo,
risk=nriskcemo$time, nrisk=nriskcemo$nrisk,
totalpts=206)
```

206 为 cemo 组试验的初始人数。

2.3.2 重构IPD数据

使用 getIPD 函数重建 comb 组 OS 曲线 IPD 数据，运行代码如下：

```
est_comb< getIPD (prep=pre_comb, armID=1,
tot.event=NULL)
```

```
summary (est_comb)
```

查看 est_comb 数据集可知，采用 IPDfromKM 包的 Kolmogorov_Smirnov(K-S) Test，得到 D 值为 0.053，p_value 为 0.99，预测值 est_comb 与 read_in 的临床数据 pre_comb 数据无显著性差异，可进行后续分析，代码如下：

```
est_cemo<-getIPD (prep=pre_cemo, armID=2,
tot.event=NULL)
```

同理获得 cemo 组 OS 曲线 IPD 数据，armID 定义为 2。

2.3.3 对比重构IPD数据与KN189的HR值

重构 IPD 数据与 KN189 报告的风险比 (hazard ratio, HR) 值对比分析的 R 语言运行代码如下：

```
est_IPD_comb_cemo<-rbind (est_comb$IPD,
est_cemo$IPD)
```

提取 comb、cemo 组的重构 IPD 为新数据集；

```
hr_est_IPD<-coxph (Surv(time,status)~treat,
data=est_IPD_comb_cemo)
```

计算重构 IPD 数据中 comb、cemo 组的 HR 值；

```
exp (coef (hr_est_IPD))
```

运行并提取 HR 值为 0.57，与 KN189 报告中数值 0.56 相近。

2.4 绘制重构IPD数据与原read_in曲线对比检验

使用 plot 绘制 comb 组与 cemo 组重构 IPD 数据与 read_in 的临床数据比较图，图 3 显示两组

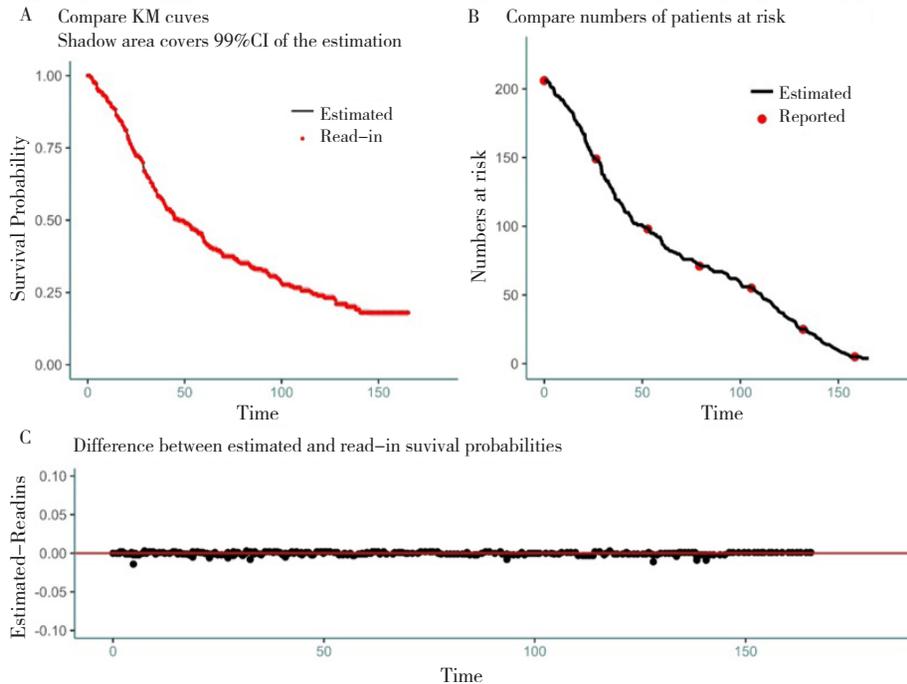


图3 cemo组重构IPD数据OS曲线与read-in KN189数据OS曲线对比

Figure 3. Comparison of OS curves of reconstructed IPD data and read-in KN189 data in cemo group

注：A. OS曲线对比图；B. 对应周期的时间风险人数对比；C. 生存概率差异。

数据重合度较高。若重构 IPD 与 read_in 曲线对比图可见明显差异，则需从第 1 步开始重复以上开始步骤，并检查原始数据，具体代码如下：

```
plot(est_comb)
```

Estimated 为重构 IPD 数据，Reported 为 read_in KN189 临床数据；

```
plot(est_cemo)
```

如图 3 所示，图 3-A 为基于重构 IPD 数据的 OS 曲线与数字化读取的 KN189 临床试验的 OS 曲线的视觉对比图。

将两组 IPD 数据分别加上 treat 列并分别赋值为 1、2，然后将 comb 与 cemo 组的重构 IPD 数据整合为数据框（表 3），具体代码如下：

```
IPD_comb_cemo_os
```

```
surcombIPD<- transform (est_comb$IPD, treat=1)
```

将治疗方案 comb 组定义为 1；

```
surcemoIPD<- transform (est_cemo$IPD, treat=2)
```

将治疗方案 cemo 组定义为 2；

```
IPD_comb_cemo_os<- rbind (surcombIPD, surcemoIPD)
```

重复以上 2.2 至 2.4 步骤，得到 OS 曲线的重构数据框 IPD_comb_cemo_os。

表3 OS曲线重构IPD数据

Table 3. Reconstructed IPD data of OS curve

患者编号	生存时间 (周)	生存状态 [#]	治疗方案 [*]
1	1.8	1	1
2	1.8	1	1
3	2.4	1	1
4	2.4	1	1
5	2.4	1	1
6	3.1	1	1
.....			
611	155.9	0	2
612	161.6	0	2
613	165.5	0	2
614	165.5	0	2
615	165.5	0	2
616	165.5	0	2

注：[#]生存状态为以每例患者为单位，在观察时刻出现结局事件的状态，“1”为发生结局事件，“0”为未发生结局事件；^{*}治疗方案中1为comb组，2为cemo组。

3 生存曲线参数分布拟合与评估

定义待拟合模型名称的向量，同时考虑指数、威布尔、伽玛、对数正态、对数逻辑及广义伽玛

多个模型，运行代码如下：

```
mods1<-c ("exponential", "weibull", "gamma",
"lognormal", "loglogistic", "gengamma")
```

定义模型公式，采用 Surv() 函数，以数据框中分类变量 treat 作为协变量分析，具体代码如下：

```
formula<- Surv (time, status)~ as.factor (treat)
```

执行批量生存模型分析，采用最大似然估计（maximum likelihood estimates, MLE），使用 flexsurv 包 fit.models() 函数创建 survHE 对象 KN189case_mle，其中存储了所考虑的多个参数模型的拟合结果，代码如下：

```
KN189case_mle<-fit.models (formula=formula,
data=IPD_comb_cemo_os, distr=modsl, method="mle")
```

```
KN189case_mle$model.fitting
```

显示批量拟合的结果 AIC、BIC、DIC 值。

基于 AIC、BIC 较小者模型拟合佳的原则，本例结果显示 AIC、BIC 值最小的拟合为 loglogistic，即 loglogistic 模型拟合度最优。同样步骤进行 MLE 批量拟合 IPD_comb_cemo_pfs，结果也显示 loglogistic 拟合模型的 AIC、BIC 值最小，

拟合度最优。

4 构建生存、效用、费用模型

4.1 定义模型

调用 R 软件 hesim 包的 hesim_data 函数^[13]，用以标准化后续生存效用模型的数据准备，代码如下：

```
hesim_dat<- hesim_data (strategies =data.
table (strategy_id=1 : 2, strategy_name=c ("comb",
"cemo")),
```

```
patients=data.table (patient_id=1, grp_id=1),
states =data. table (state_id=1 : 2, state_name=c
("Stable", "Progression")))
```

分别定义分区生存模型中治疗方案 strategies、患者类型 patients、试验疾病状态 states，死亡状态不用列出。

```
labs<- get_label (hesim_dat)
```

4.2 建立生存模型

将 IPD_comb_cemo_os、IPD_comb_cemo_pfs 数据合并，并整理为数据框 surv_est_data，整理后的数据结构见表 4。

表4 comb与cemo组OS及PFS曲线的重构患者数据

Table 4. Reconstructed patient data of OS and PFS curves in comb group and cemo group

PFS曲线			OS曲线		
生存时间（周）	生存状态 [#]	治疗方案 [*]	生存时间（周）	生存状态 [#]	治疗方案 [*]
2.23	1	1	1.76	1	1
2.23	1	1	1.76	1	1
2.24	1	1	2.42	1	1
2.24	1	1	2.42	1	1
2.24	1	1	2.42	1	1
2.79	1	1	3.08	1	1
.....				
113.08	1	2	155.90	0	2
114.62	0	2	161.62	0	2
123.64	0	2	165.46	0	2
134.20	1	2	165.46	0	2
137.06	0	2	165.46	0	2
138.60	1	2	165.46	0	2

注：[#]生存状态为以每例患者为单位，在观察时刻出现结局事件的状态，“1”为发生结局事件，“0”为未发生结局事件；^{*}治疗方案中1为comb组，2为cemo组。

采用 flessurvreg() 函数，基于重构 IPD 数据集 surv_est_data，分别拟合 OS 及 PFS 生存模型，代码如下：

```
flex_fit_os<- flessurvreg (Surv (os_time, os_
status) ~strategy_name, data=surv_est_data,
dist= "llogis" )
```

选择步骤 3 中 pfs 及 os 曲线最优拟合模型 llogis:

```
flex_fit_pfs<- flexsurvreg (Surv(pfs_time, pfs_
status)~strategy_name, data=susurv_est_data,
dist="llogis")
```

综合 flex_fit_os 及 flex_fit_pfs 模型为 psfit_wei, 代码如下:

```
psfit_wei<- flexsurvreg_list (flex_fit_pfs, flex_
fit_os)
```

包含了 3 个状态分别的生存拟合结果对象。

建立生存模型 survmods, 调用 hesim 程序包的 create_PsmCurves() 函数。在其参数设定中, 明确 bootstrap 为确定生成结果不确定性的方法, n 为 bootstrap 的样本数, 运行代码如下:

```
n_samples<- 300
```

设定生存模型中将要模拟的样本数量;

```
surv_input_data<- expand(hesim_dat,
by=c("strategies", "patients") )
```

从 hesim_dat 中指定参数进行模拟;

```
survmods<- create_PsmCurves (psfit_wei, n=n_
samples, input_data=surv_input_data,
uncertainty="bootstrap", est_data="surv_est_
data")
```

input_data: hesim 数据框架, 包含了需要生成生存函数的所有患者和策略组合形式; n 为用于 bootstrap 的样本数; uncertainty 用于确定生成结果的不确定性的方法, 本例使用 bootstrap 方法; est_data: hesim 数据集, 类似于 IPD 生存状态与时间数据及其他如年龄等所有额外信息。

4.3 建立效用模型

疾病状态的效用值采用文献数据^[21], 即 pfs=0.815、pd=0.321, 其效用值参数的变化范围: DSA 在 +15% 范围, PSA 为 beta 分布。据此定义状态效用值数据框 (表 5), 运行代码如下:

```
utility_tbl<- stateval_tbl (utility_data,
dist="beta")
```

```
utilitymod<- create_StateVals (utility_tbl, n=n_
samples, hesim_data=hesim_dat)
```

表5 不同状态效用值

Table 5. Utility values for different states

状态 [#]	效用平均值	效用值参数1	效用值参数2
1	0.82	2.8	0.63
2	0.32	29.7	62.88

注: #状态中1为无进展, 2为进展。

4.4 建立费用模型

根据 KN189 临床设计方案, 针对使用药品费用 (包括疾病进展后使用二线治疗方案药物、不同方案治疗比例)、检查检验的周期计划及费用、不良反应发生率及相应治疗费用、临终关怀费用等, 采用药智网及北京市医疗费用收费标准进行综合计算, 得出 comb、cemo 组各自总体费用的均值 (表 6), 运行代码如下:

```
cost_strategy_id
```

```
cost_tbl<- stateval_tbl (cost_strategy_id,
dist="fixed")
```

```
costmod<- create_StateVals (cost_tbl, n=n_
samples, hesim_data=hesim_dat)
```

表6 不同治疗方案费用

Table 6. The cost of different plans

治疗方案 [#]	估算不同方案费用 (元)
1	193 263
2	135 950

注: #治疗方案中1为comb组, 2为cemo组。

5 实例化PSM并计算质量调整生命年

5.1 定义模型

生存模型 survmods 与后续的效用模型、费用模型汇总进行 PSM 的实例化, 并进行时间外推, 分析代码如下:

```
psm<- Psm$new (survival_models=survmods,
utility_model=utilitymod, cost_models=list
(medical=costmod))
```

```
times<- seq (0, 1000, by=50)
```

定义生存模型实例化外推时间的范围为 20 年, 与前数据统一以周为单位, 将时间范围带入 psm 进行生存预测;

```
psm$sim_survival (t=times)
```

将定义的外推时间输入 psm 进行生存预测。

5.2 绘制外推生存曲线图

基于以上所定义的外推时间范围 times 内, 将实例化 PSM 中的生存概率绘制生存曲线, 结果见图 4, 分析代码如下:

```
labs$curve<- c("PSF_Curve"=1, "OS_Curve"=2)
```

```
autoplot (psm$survival_, lables=labs, ci=TRUE,
ci_style="ribbon",
```

```
scale_x_continuous (breaks=seq (0, max (times), 50))
```

```
psm$sim_stateprobs())
```

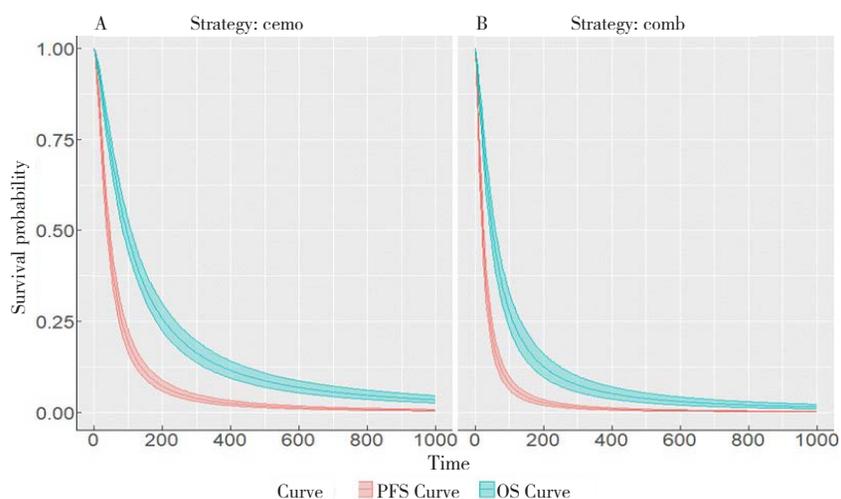


图4 外推时间内的PFS和OS曲线
Figure 4. PFS and OS curves over time

注：A. cemo组对比图；B. comb组对比图。

整理状态概率数据框，用于后续视图表现；
`stateprobs <- psm$stateprobs_[sample==2&patient_id==1]`
 从 psm 对象中任选一个子集，选择 sample=2 且 patient_id=1，将其状态概率定义给到 stateprobs；
`stateprobs[, state:=factor(state_id, state, levels=rev(unique(state_id)))]`
 在 stateprobs 数据框中增加 state 列，将 state_id 转换为因子类型的变量赋值给 state 列，并指定其级别为唯一值且逆序排列；

同理增加 strategy 列如下：
`stateprobs[, strategy:=factor(strategy_id, labels=c(hesim_dat$strategies$strategy_name))]`
 基于整理的 stateprobs 绘制状态概率曲线，见图 5，分析代码如下：
`ggplot(stateprobs[strategy_id %in% c(1:2)])`
 指定数据来源 stateprobs 中 strategy_id 列数据绘图；
`aes(x=t, y=prob, fill=state, group=state)+`
 定义 x、y 轴分，根据 state 列不同值着色分组；

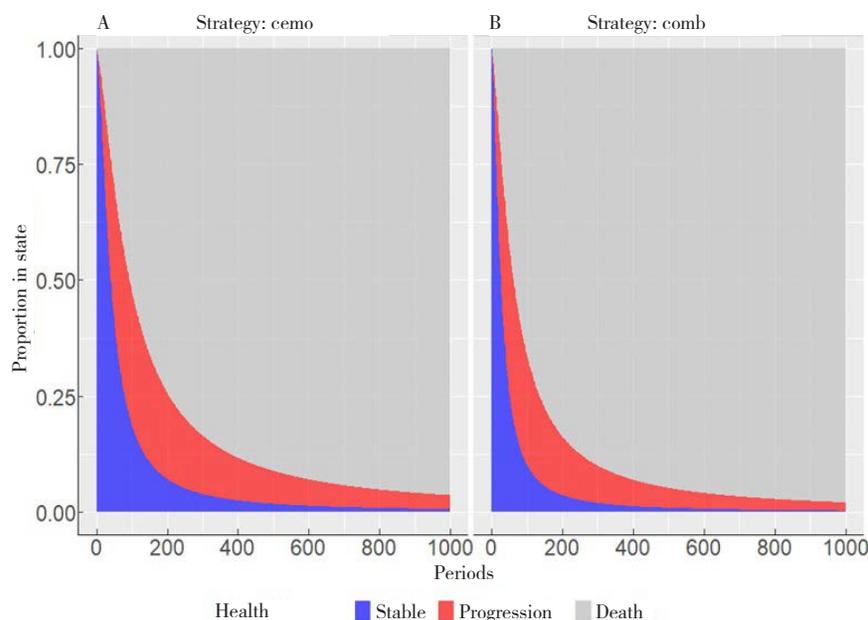


图5 外推时间内的状态曲线
Figure 5. State curves over time

注：A. cemo组对比图；B. comb组对比图。

```
geom_area(alpha=.65), facet_wrap(~strategy)
将图形分面，每个分面表示不同的策略
(strategy);
abs(x="Periods", y="Proportion in state")
设定 x、y 轴的文字标识;
scale_fill_manual(name="Health state")
设置图例的标题;
values=c("gray", "red", "blue")
设置填充图形的颜色映射;
lables=c("Death", rev(hesim_dat$states$state_
name)))+
标识每个状态的文字内容;
scale_x_continuous(breaks=seq(0, max(times), 50)+
设置 x 轴的刻度标签;
guides(fill=guide_legend(reverse=T, nrow=2,
byrow=TRUE))+
调整图例以逆序的方式且排 2 行;
theme(legend.position="bottom")
设置图例的位置在底部。
```

5.3 模拟成本与质量调整生命年

基于 5.2 中的状态概率进行数值积分模拟贴现成本与质量调整生命年 (quality-adjusted life

years, QALYs), 选择 5% 进行贴现, 分析代码如下:

```
psm$sim_costs(dr=0.05)
```

运行后结果见图 6。

```
psm$sims_qalys(dr=0.05)
```

运行后结果见图 7。

6 视图辅助决策的成本效用分析

6.1 创建模拟数据

调用 hesim 包的 cea() 和 cea_pw() 函数分别执行单一或成对的两种治疗策略的成本效益分析, 其运行结果对象中包含有每个 PSA 样本的平均成本与 QALYs。先用 summarize() 方法创建一个具有平均值的 hesim::ce 对象, 并将其赋值给变量 ce_sim, 该对象包括了在概率敏感分析 PSA 中使用的模拟结果的摘要统计信息, 分析代码如下:

```
ce_sim<-psm$summarize()
```

```
wtp<-seq(from=0, to=20000, by=2000)
```

定义支付意愿阈值 wtp 的范围, 以此进行 CUA 分析;

```
cea_out<-cea(ce_sim, dr_qalys=.05,
```

```
dr_costs=.05, k=wtp)
```

将 cea() 函数分析结果赋值给变量 cea_out;

```
> psm$costs_
  sample strategy_id patient_id grp_id state_id  dr  category  costs
  <num>    <int>    <int>    <int>    <int> <num>    <char>    <num>
1:      1          1          1          1          1 0.05 Category 1 3514455.4
2:      1          1          1          1          2 0.05 Category 1 182947.4
3:      1          2          1          1          1 0.05 Category 1 5173299.5
4:      1          2          1          1          2 0.05 Category 1 252119.0
5:      2          1          1          1          1 0.05 Category 1 3546765.5
---
1196:  299          2          1          1          2 0.05 Category 1 264999.3
1197:  300          1          1          1          1 0.05 Category 1 3532354.3
1198:  300          1          1          1          2 0.05 Category 1 182856.8
1199:  300          2          1          1          1 0.05 Category 1 5179787.6
1200:  300          2          1          1          2 0.05 Category 1 249971.4
```

图6 基于PSM以5%的折现率模拟的贴现成本

Figure 6. Discounted cost simulated based on PSM at a 5% discount rate

```
> psm$qalys_
  sample strategy_id patient_id grp_id state_id  dr  qalys  lys
  <num>    <int>    <int>    <int>    <int> <num>    <num>    <num>
1:      1          1          1          1          1 0.05 20.5712264 25.851126
2:      1          1          1          1          2 0.05 0.4640832 1.345698
3:      1          2          1          1          1 0.05 21.3009491 26.768142
4:      1          2          1          1          2 0.05 0.4498879 1.304536
5:      2          1          1          1          1 0.05 24.6411727 26.088788
---
1196:  299          2          1          1          2 0.05 0.5454191 1.371183
1197:  300          1          1          1          1 0.05 25.9816560 25.982784
1198:  300          1          1          1          2 0.05 0.3684051 1.345032
1199:  300          2          1          1          1 0.05 26.8005501 26.801714
1200:  300          2          1          1          2 0.05 0.3542697 1.293424
```

图7 基于PSM以5%的折现率模拟的贴现效用

Figure 7. Discounted utilities simulated based on PSM at a 5% discount rate

```
cea_pw_out<- cea_pw (ce_sim, comparator=1,
dr_qalys=.05, dr_costs=.05, k=wtp)
```

cea_pw() 函数计算人群加权增量成本-效果比 (population-weighted incremental cost-effectiveness ratio, PWICER), comparator=2 即将 strategy_id=2 的 cemo 组为基准策略分析偏好权重, K 为支付意愿阈值。

6.2 绘制成本效果可接受曲线

从 cea_out 中提取 strategy_id=1 的 comb 组成本效果数据, 并绘制单组的成本效果可接受曲线, 分析代码如下:

```
strategy_1_data<- cea_out$mce [strategy_id==1]
cea_ac_plot<- ggplot (data=strategy_1_data,
aes(x=strategy_1_data$k, y=strategy_1_data$prob,
group=1))+
geom_line (linewidth=1), xlab ("k")+ylab
("prob")+
ggtitle ("Cost-Effectiveness Acceptability
Curve"), theme_minimal()
```

提取 strategy_id=2 的 cemo 组数据, 将两组数据合并后, 同时绘制 cemo 与 comb 组的成本效果可接受曲线:

```
strategy_2_data<- cea_out$mce [strategy_id==2]
combined_data<- rbind (strategy_1_data,
strategy_2_data)
colors<- c ("blue", "red")
labels<- c ("Strategy 1", "Strategy 2")
cea_ac_plot<- ggplot (data=combined_data, aes
(x=k, y=prob, color=factor (strategy_id))) +
geom_line (linewidth=1), xlab ("Willingness to
pay"), ylab ("Probability Acceptable")+
scale_color_manual (values=colors,
lables=lables),
guides (color=guide_legend (title="Strategy"))+
ggtitle ("Cost-Effectiveness Acceptability
Curve"), theme_minimal()
print(cea_ac_plot)
```

由图 8 可知, 当支付意愿值超过 4 800 元时, comb 组具有成本效益的可能性开始大于 cemo 组。

6.3 绘制增量成本-效益比值散点图

绘制增量成本-效益比值 (incremental cost-effectiveness ratio, ICER) 散点图, 设定支付意愿值上下限并在图中添加 97.5% 和 2.5% 可能性的

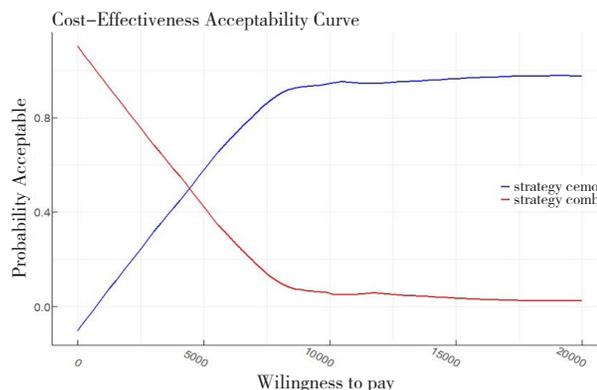


图 8 comb 组与 cemo 组成本效果可接受曲线

Figure 8. Acceptable cost-effectiveness curves of the comb group and the cemo group

正比例函数线条, 运行代码如下:

```
strategy_1<- cea_pw_out$delta [strategy_id==1, ]
提取 cea_pw_out 数据集中 comb 组的数据;
strategy_1$icer<- strategy_1$ic / strategy_1$ie
计算 comb 组的 ICER;
wtp_2<- quantile (strategy_1$icer, probs=0.975)
wtp_3<- quantile (strategy_1$icer, probs=0.025)
cea_sensitivity_plot<- ggplot (data=cea_pw_
out$delta,
aes (x=ie, y=ic, color=strategy_id, size=3,
alpha=0.8))+
geom_point (color="red", size=3, alpha=0.8)+
geom_abline (slope=wtp_2, intercept=0,
color="blue", linetype="dashed")+
geom_abline (slope=wtp_3, intercept=0,
color="red", linetype="dashed")+
geom_text (data=data.frame (x=0.2, y=3 000,
lable=paste0 ("97.5% ICER=", round (wtp_2,2))),
geom_text (data=data.frame (x=1.8, y=3 000,
lable=paste0 ("2.5%ICER=", round (wtp_3,2))),
xlab ("Incremental Effectiveness"),
ylab("Incremental Cost")+
ggtitle ("Sensitivity Analysis"), xlim (0, 1.25),
ylim (1650000, 1800000))+
theme_minimal()
print (cea_sensitivity_plot)
```

由图 9 可知, 当意愿支付阈值为 14 312 时, comb 组干预方案具有成本-效益的可能性为 97.5%; 当意愿支付阈值为 3 123 时, comb 组干预方案具有成本-效益的可能性仅为 2.5%。

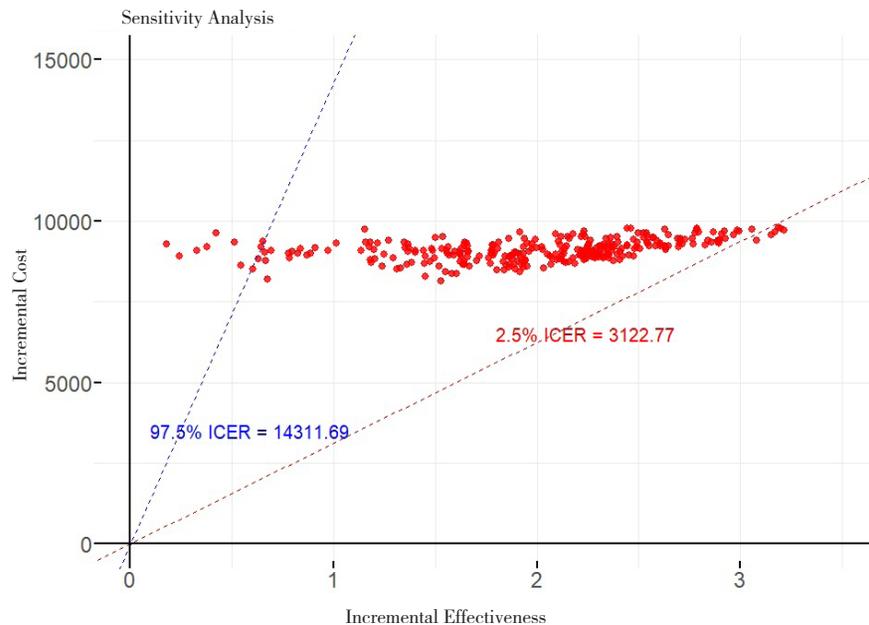


图9 comb组Bootstrap自助法抽样300次后得到的ICER散点图

Figure 9. Scatter plot of ICER obtained after 300 iterations of Bootstrap resampling in the comb group

7 讨论

本研究展示了 R 语言在成本效用分析中的高效性，具体表现在：一是代码透明性，分析过程完全通过代码记录，确保模型运行的透明性；二是计算可重复性，仅需重新运行代码即可重复整个计算过程并更新结果；三是错误识别与调整便捷性，模型设计及参数设定的错误或假设调整易于发现与更正；四是结果输出直接性，PE 评价结果可直接通过 R 代码生成；五是敏感分析，可通过扩大分析的范围评估关键参数和假设变化对模型输出的影响，从而捕捉不确定性、识别关键参数及优化模型；六是结果输出直接性，PE 评价结果可直接通过 R 代码生成。此外，还具有模型的灵活性，hesim 程序包^[13]为健康经济学模拟提供了一个灵活的框架，除 PSM 外，还提供马尔科夫模型（Markov models）、状态转换模型（state-transition models, STM）、队列状态转换模型（cohort state-transition models）和个体状态转换模型（individual state-transition models）。上述模型均支持对成本和效用的模拟分析，适用于多种健康状态和时间点的数据。

PSM 作为决策支持工具，在本研究中展现了其优势与局限性。通过 GetDataW 软件提取

KN189 试验的 PFS 和 OS 数据，为生存分析提供了基础，避免了获取临床试验 IPD 的需求。然而，PSM 的应用依赖于临床试验文献对生存曲线的详细报告，若需分析特定亚组或多状态患者的时间相关数据，而文献未提供必要信息时，分析可能受限。此外，PSM 能准确模拟疾病事件，但在随访时间超出临床试验报告时，需通过参数法外推 PFS 和 OS 曲线，这一过程对于晚期和转移性肿瘤药物的成本效用分析至关重要，但需要注意 PSM 方法本身限制了敏感性分析的范围，当长期的总生存期外推存在不确定性时，局限性就会显现。因此，建议将 PSM 与 STM 一起使用，以支持对其推断的评估^[4]。研究者需仔细检查外推拟合结果，利用个体临床事件的试验数据、外部数据和专家意见综合评估 PSM 和 STM 的生存预测可信度，避免出现不合理的统计现象^[13]，后续进一步通过实例进行研究和说明。采用 R 语言进行肿瘤治疗药物的成本效用分析要求研究者具备临床知识、数理统计能力及编程经验，研究者需在临床方案设计理解、参数设定、模型假设及验证等方面谨慎设定与检查，以确保分析结果的真实性和可靠性。

本研究聚焦于使用 hesim 建立模型及 PSM 辅助决策的视图功能，未能详尽阐述 R 软件及相关

程序包安装、PFS 和 OS 曲线的数字化提取、参数分布的拟合与评估、生存分析的完整性及费用模型的数据来源和计算方法等内容,感兴趣的读者可参考文献^[10-17],以获取更多关于上述内容的详细信息。本研究为了突出 PSM 模型建立,采用简单数值演示费用及效用模型,若需进行复杂的费用及效用模型分析,如疾病进展后多线治疗方案的比例、不良反应发生率及治疗费用等,则需单独建立模型,然后并入第 5 步的实例化 PSM,并进行概率敏感性分析。

参考文献

- 1 Woods BS, Sideris E, Palmer SJ, et al. NICE DSU technical support document 19: partitioned survival analysis for decision modelling in health care: a critical review[EB/OL]. (2017-06-02). <https://www.sheffield.ac.uk/sites/default/files/2022-02/TSD19-Partitioned-Survival-Analysis-final-report.pdf>
- 2 Bullement A, Cranmer HL, Shields GE. A review of recent decision-analytic models used to evaluate the economic value of cancer treatments[J]. *Appl Health Econ Health Policy*, 2019, 17(6): 771-780. DOI: 10.1007/s40258-019-00513-3.
- 3 孟蕊, 芮明军, 王欣恬, 等. 不同决策分析模型在抗肿瘤药物经济学评价中的应用——以加拿大 CADTH 为例 [J]. *中国药房*, 2021, 32(14): 1752-1757. [Meng R, Rui MJ, Wang XT, et al. Application of different decision analysis models in economic evaluation of antitumor drugs: taking CADTH in Canada as an example[J]. *Chinese Pharmacy*, 2021, 32(14): 1752-1757.] DOI: 10.6039/j.issn.1001-0408.2021.14.15.
- 4 Woods BS, Sideris E, Palmer S, et al. Partitioned survival and state transition models for healthcare decision making in oncology: where are we now?[J]. *Value Health*, 2020, 23(12): 1613-1621. DOI: 10.1016/j.jval.2020.08.2094.
- 5 Green N, Lamrock F, Naylor N, et al. Health economic evaluation using Markov models in R for Microsoft Excel users: a tutorial[J]. *Pharmacoeconomics*, 2023, 41(1): 5-19. DOI: 10.1007/s40273-022-01199-7.
- 6 邵荣杰, 唐文熙, 马爱霞. 分区生存模型在药物经济学评价中的应用 [J]. *中国卫生经济*, 2019, 38(9): 60-63. [Shao RJ, Tang WX, Ma AX. The partitioned survival model applied in pharmacoeconomic evaluation[J]. *Chinese Health Economics*, 2019, 38(9): 60-63.] DOI: 10.7664/CHE20190916.
- 7 曾小慧, 彭六保, 谭重庆, 等. 药物经济学评价中的分区生存模型 [J]. *中国新药与临床杂志*, 2020, 39(8): 504-507. [Zeng XH, Peng LB, Tan CQ, et al. Partitioned survival model in pharmacoeconomics[J]. *Chinese Journal of New Drugs and Clinical Remedies*, 2020, 39(8): 504-507.] DOI: 10.14109/j.cnki.xyylc.2020.08.12.
- 8 刘新义, 谭重庆, 曾小慧, 等. 分区生存模型在药物经济学评价中的应用简介及实例解析 [J]. *中国现代应用药学*, 2019, 36(24): 3090-3093. [Liu XY, Tan CQ, Zeng XH, et al. Introduction of partitioned survival model in pharmacoeconomics evaluation and case analysis[J]. *Chinese Journal of Modern Applied Pharmacy*, 2019, 36(24): 3090-3093.] DOI: 10.13748/j.cnki.issn1007-7693.2019.24.018.
- 9 桂裕亮, 韩晟, 翁鸿, 等. 应用 TreeAge Pro 软件实现基于 Markov 模型的成本-效果分析 [J]. *中国循证医学杂志*, 2018, 18(1): 116-120. [Gui YL, Han S, Weng H, et al. Cost-effectiveness analysis based on Markov model using TreeAge Pro software[J]. *Chinese Journal of Evidence-Based Medicine*, 2018, 18(1): 116-120.] DOI: 10.7507/1672-2531.201707008.
- 10 Filipović-Pierucci A, Zarca K, Durand-Zaleski I. Markov models for health economic evaluations in R with the heemod package[J]. *Value Health*, 2016, 19(7): 369. DOI: 10.1016/j.jval.2016.09.133.
- 11 母立峰, 苏亚霞, 宋金春. 基于 R 语言的药物经济学研究方法综述 [J]. *药物流行病学杂志*, 2019, 28(7): 472-476. [Mu LF, Su YX, Song JC. A review of the research method of pharmacoeconomics based on R language[J]. *Chinese Journal of Pharmacoepidemiology*, 2019, 28(7): 472-476.] DOI: 10.19960/j.cnki.issn1005-0698.2019.07.013.
- 12 Jackson CH. flexsurv: a platform for parametric survival modeling in R[J]. *J Stat Softw*, 2016, 70: i08. DOI: 10.18637/jss.v070.i08.
- 13 Incerti D, Jansen JP. hesim: health economic simulation modeling and decision analysis[EB/OL]. (2021-03-09) [2023-06-30]. <https://doi.org/10.48550/arXiv.2102.09437>.
- 14 Moore DF. *Applied survival analysis using R*[M]. Cham: Springer, 2016.
- 15 Baio G. survHE: survival analysis for health economic evaluation and cost-effectiveness modeling[J]. *J Stat Softw*,

- 2020, 95(14): 1–47. DOI: [10.18637/jss.v095.i14](https://doi.org/10.18637/jss.v095.i14).
- 16 石丰豪, 商叶, 芮明军, 等. R 语言 survHE 程序包在卫生经济学评价中的应用[J]. 中国卫生经济, 2020, 39(9): 9–14. [Shi FH, Shang Y, Rui MJ, et al. Application of survHE package of R for health economic evaluation[J]. China Health Economics, 2020, 39(9): 9–14.] DOI: [10.7664/CHE20200902](https://doi.org/10.7664/CHE20200902).
- 17 Baio G, Berardi A, Heath A. Bayesian cost-effectiveness analysis with the R package BCEA[M]. New York: Springer, 2017.
- 18 Ben ÂJ, van Dongen JM, El Alili M, et al. Conducting trial-based economic evaluations using R: a tutorial[J]. Pharmacoeconomics. 2023, 41(11): 1403–1413. DOI: [10.1007/s40273-023-01301-7](https://doi.org/10.1007/s40273-023-01301-7).
- 19 Rodríguez-Abreu D, Powell SF, Hochmair MJ, et al. Pemetrexed plus platinum with or without pembrolizumab in patients with previously untreated metastatic nonsquamous NSCLC: protocol-specified final analysis from KEYNOTE-189[J]. An Oncol, 2021, 32(7): 881–895. DOI: [10.1016/j.annonc.2021.04.008](https://doi.org/10.1016/j.annonc.2021.04.008).
- 20 Gadgeel S, Rodríguez-Abreu D, Speranza G, et al. Updated analysis from KEYNOTE-189: pembrolizumab or placebo plus pemetrexed and platinum for previously untreated metastatic nonsquamous non-small-cell lung cancer[J]. J Clin Oncol, 2020, 38(14): 1505–1517. DOI: [10.1200/JCO.19.03136](https://doi.org/10.1200/JCO.19.03136).
- 21 徐赫, 马爱霞. 基于分区生存模型的帕博利珠单抗单药与化疗一线治疗 PD-L1 肿瘤比例分数不同的非小细胞肺癌的成本-效果分析[J]. 中国医院药学杂志, 2020, 40(23): 2468–2473. [Xu H, Ma AX. Cost-effectiveness analysis of pembrolizumab versus chemotherapy as first-line treatment in non-small cell lung cancer with different PD-L1 expression levels based on partitioned survival model[J]. Chinese Journal of Hospital Pharmacy, 2020, 40(23): 2468–2473.] DOI: [10.13286/j.1001-5213.2020.23.14](https://doi.org/10.13286/j.1001-5213.2020.23.14).

收稿日期: 2024 年 01 月 05 日 修回日期: 2024 年 04 月 12 日
本文编辑: 张 苗 黄 笛

引用本文: 刘家一, 李伟. R 语言实例解析肿瘤治疗药物经济学评价的分区生存模型[J]. 数理医药学杂志, 2024, 37(4): 240–251. DOI: [10.12173/j.issn.1004-4337.202401034](https://doi.org/10.12173/j.issn.1004-4337.202401034).
Liu JY, Li W. Partitioned survival model of pharmacoeconomics evaluation of oncology therapy in R language[J]. Journal of Mathematical Medicine, 2024, 37(4): 240–251. DOI: [10.12173/j.issn.1004-4337.202401034](https://doi.org/10.12173/j.issn.1004-4337.202401034).