

现时法统计模型在全人群肿瘤登记生存分析中的应用

李卓颖^{1,2} 蒋宇飞¹ 谭玉婷^{1,2} 项永兵^{1,2△}

(¹上海交通大学医学院附属仁济医院上海市肿瘤研究所癌基因及相关基因国家重点实验室 上海 200032;

²上海市肿瘤研究所流行病学研究室 上海 200032)

【摘要】 目的 系统介绍现时法统计模型的基本原理及其在全人群肿瘤登记生存数据分析中的应用。**方法** 首先介绍常规的现时法,并在此基础上对现时法统计模型的基本原理进行了详细描述,最后以实例说明如何在R软件中应用现时法统计模型进行相对生存率的计算和趋势检验。**结果** 利用上海市区男、女性胃癌1972—1986年生存随访数据作为实例,分别采用队列法、完全法、现时法和现时法统计模型估计1977—1981年胃癌5年相对生存率,并对相对生存率在1978—1980年、1981—1983年和1984—1986年3个观察时期间的趋势变化进行统计学检验。生存率估计结果显示,4种方法估计的5年相对生存率分别为:男性12.77%、13.76%、14.27%和16.41%;女性11.51%、12.04%、12.35%和14.74%。现时法统计模型估算的结果与后来实际观测到的胃癌5年相对生存率(男性:16.46%,女性:14.23%)较接近。趋势检验结果提示,胃癌5年相对生存率在3个观察时期间的趋势变化有统计学意义。**结论** 现时法统计模型可在保证精确度的前提下,提高生存率估计的时效性,还可用于评估相关协变量对生存率的影响。国内肿瘤登记处在分析全人群肿瘤生存数据时,可以根据实际情况加以应用和验证。

【关键词】 肿瘤登记; 生存分析; 相对生存率; 现时法统计模型

【中图分类号】 R181.2+2 **【文献标志码】** A **doi:** 10.3969/j.issn.1672-8467.2021.03.015

Application of model-based period analysis in estimating survival rates for population-based cancer registration data

LI Zhuo-ying^{1,2}, JIANG Yu-fei¹, TAN Yu-ting^{1,2}, XIANG Yong-bing^{1,2△}

(¹State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200032, China; ²Department of Epidemiology,

Shanghai Cancer Institute, Shanghai 200032, China)

【Abstract】 Objective To systematically introduce the theory and computational methods of the model-based period analysis in estimating survival rates for the population-based cancer registration survival data. **Methods** The conventional period analysis was firstly introduced, and then the model-based period analysis was described in detail. In addition, two examples were given to illustrate how the model-based period analysis may be used in the estimation and trend test of relative survival rates in R software. **Results** The survival data of male and female gastric cancer in urban Shanghai from 1972 to 1986 were used in the empirical examples. The 5-year relative survival rates of gastric cancer in urban Shanghai during 1977 to 1981 were estimated by cohort method, complete method, period method and model-based period analysis. A trend test was conducted for the relative survival rates over the periods of 1978–1980, 1981–1983 and 1984–1986. The estimates of 5-year relative survival rates for the above methods were 12.77%, 13.76%, 14.27% and 16.41% for men, and 11.51%, 12.04%, 12.35% and 14.74% for women,

国家重点研发计划(2016YFC1302503)

[△]Corresponding author E-mail: ybxian@shsci.org

网络首发时间:2021-05-12 10:34:51 网络首发地址:https://kns.cnki.net/kcms/detail/31.1885.R.20210510.2109.022.html

respectively. The results of the model-based period analysis were much closer to what we observed later (16.46% for men and 14.23% for women). The trend test indicated that the improvement of the relative survival rates of gastric cancer were statistically significant during the three periods. **Conclusion** The model-based period analysis improves the up-to-dateness of survival estimation on the premise of precision. It can also be used to evaluate the effect of relevant covariates on the survival rates of cancer patients. The method could be applied and validated in the analysis of population-based cancer registration survival data.

【Key words】 cancer registry; survival analysis; relative survival rate; model-based period analysis

* This work was supported by the National Key R&D Program of China (2016YFC1302503).

肿瘤患者的生存率是评估一个地区或国家肿瘤综合诊治措施效果的重要指标,一般可由临床随访研究、医院肿瘤登记管理数据和人群肿瘤登记资料计算而得。其中,人群肿瘤登记数据由于覆盖一个固定人群,计算得到的肿瘤生存率代表性较好,能够真实反映一个地区内全部人群的肿瘤生存情况^[1]。

生存率计算往往要回顾和整理过去一段时间的历史资料,关键问题是选择研究的观察时间窗。两种传统的生存率计算方法需要长时间的完整随访资料,纳入观察的对象多来源于多年前诊断的病例,无法及时反映肿瘤诊治水平提高所带来的生存率改善。队列法(cohort method)仅纳入完全随访的病例资料,时效性和精确度有一定局限性;完全法(complete method)在队列法的基础上加入了尚未完成随访的右截尾数据,估计的精确度有所提高,但时效性仍较差。Brenner等^[2]在1996年提出现时法(period analysis),仅纳入感兴趣时间段内观察到的生存资料,包括在此时间段内完成随访的左截尾数据和新近诊断的右截尾数据。现时法排除了多年前的随访资料,提高了生存率估计的时效性,但由于纳入计算的有效病例数也随之减少,故生存率估计的标准误有所增大,精确度不如完全法。因此,实际应用中常面临时效性和精确度的权衡,两者往往难以兼顾。2006年,Brenner等^[3]提出了现时法统计模型(model-based period analysis),在常规现时法的基础上通过拟合广义线性模型来考虑观察时期对生存率的影响。该模型可在充分利用现有数据、保证精确度的前提下,提高生存率估计的时效性,还可以用于评估其他预后相关因素对生存率的影响。目前,现时法统计模型在欧美国家的肿瘤登记中已得到一定范围的应用^[4-10],但国内至今尚无相关的研究报道。

鉴于此,本研究将系统介绍现时法统计模型的基本原理,并以上海市肿瘤登记处收集整理的胃癌历史生存资料为例,简述其在R软件中如何实现。

资料和方法

基本原理 人群肿瘤登记数据最常用的生存率统计指标有观察生存率(observed survival rate, OSR)和相对生存率(relative survival rate, RSR)^[11]。生存率的计算通常采用寿命表法,例如5年生存率由1~5年的一年区间生存概率累乘而得,一年区间生存概率由观察区间内实际的观察死亡数(d)、期望死亡数(e)和区间内的随访人数(l)计算而来^[12]。队列法、完全法、现时法的区别仅在于用于计算一年区间生存概率的数据在诊断时间和随访时间上的差异。

以估计患者的5年RSR为例,现时法统计模型^[3]是在常规现时法^[2]算法的基础上,以随访年份 i (分类变量,计算5年生存率时 $1 \leq i \leq 5$)和观察时期 j (连续型变量,第1个观察时期 $j=0$,往后逐渐加1)为自变量,以每个观察区间内的实际观察死亡数(d_{ij})为因变量,拟合一个广义线性模型。误差结构取Poisson分布,连接函数定义为 $\ln(\mu_{ij} - d_{ij}^*)$, μ_{ij} 为基于模型的估计死亡数, $d_{ij}^* = -\left(l_{ij} - \frac{d_{ij}}{2}\right) \times \ln\left(\frac{l_{ij} - e_{ij}}{l_{ij}}\right)$, e_{ij} 和 l_{ij} 分别表示期望死亡数(来自当地人群寿命表)和区间内随访人数。模型拟合时把有效观察人数的自然对数 $\ln\left(l_{ij} - \frac{d_{ij}}{2}\right)$ 作为偏移项(offset),省略截距项。下述模型可以用于描述超额

死亡数与随访年份及观察时期的关系:

$$\ln(\mu_{i,j} - d_{i,j}^*) = \ln\left(l_{i,j} - \frac{d_{i,j}}{2}\right) + \sum_i \alpha_i + j \times \beta$$

其中, α_i 表示第 i 个随访年份的回归系数, β 表示每增加 1 个观察时期的回归系数。则每一年的区间生存概率 $r_{i,j} = \exp(-\exp(\alpha_i + j \times \beta))$; 第 j 个观察时期的 5 年生存率 $R_j = \prod_{i=1}^5 r_{i,j} = \prod_{i=1}^5 \exp(-\exp(\alpha_i + j \times \beta))$ 。上述公式也适用于观察生存率的计算, 只需将 $d_{i,j}^*$ 设为 0 即可。

生存率标准误 (standard error, SE) 的估计可使用 delta 法^[13], 设 $\text{var}(\alpha_i)$ 、 $\text{var}(\beta)$ 、 $\text{cov}(\alpha_i, \alpha_k)$ 和 $\text{cov}(\alpha_i, \beta)$ 分别为 α_i 和 β 的方差和协方差矩阵, 则第 j 个观察时期 5 年生存率的 SE 为:

$$\begin{aligned} \text{Var}(R_j) = & \sum_{i=1}^5 (A_{i,j}^2 \times \text{var}(\alpha_i)) + B_j^2 \times \text{var}(\beta) \\ & + 2 \times \sum_{i=1}^5 (B_j^2 \times A_{i,j} \times \text{cov}(\alpha_i, \beta)) \\ & + \sum_{i=1}^5 \sum_{\substack{k=1 \\ i \neq k}}^5 (A_{i,j} \times A_{k,j} \times \text{cov}(\alpha_i, \alpha_k)) \end{aligned}$$

其中, $A_{i,j} = R_j \times \ln(r_{i,j})$, $B_j = j \times R_j \times \sum_{i=1}^5 \ln(r_{i,j})$ 。

实例分析 上海是我国最早开展城市人群肿瘤登记报告工作的地区, 积累了长期、标准、规范、完整的肿瘤登记资料^[14]。我们利用 1972—1986 年上海市区的胃癌生存资料, 以两个简单实例说明现时法统计模型计算相对生存率的基本过程以及对协变量进行统计学检验的方法。实例 1 以 1977—1981 年为感兴趣时期, 采用队列法、完全法、现时法和现时法统计模型分别计算上海市区男女性胃癌的 5 年 RSR 及其 SE, 并与之后实际观察到的生存率进行比较 (图 1)。实例 2 利用现时法统计模型检验胃癌的 5 年 RSR 在 1978—1980 年、1981—1983 年和 1984—1986 年 3 个观察时期间的趋势变化是否存在统计学意义 (图 1)。RSR 的计算采用 Ederer II 法^[11,15], 随访截止时间定义为 1986 年 12 月 31 日。所有统计分析采用 R 4.0.3 版本进行, 计算前需预先安装并调用 Brenner 等编写的 R 拓展包“PeriodR”^[16] (文献中下载地址: <https://krebsregister.saarland.de/daten-auswertungen-veroeffentlichungen/software/periodr-english/>)。实例 1 和实例 2 所采用的生存数据如图 1 所示 (完整程序见电子附加材料)。

结 果

实例 1 传统的队列法、完全法和现时法均可使用“PeriodR”拓展包实现, 该拓展包的核心是 period 函数, 其基本语句如下:

```
period (data, k, surv.m, surv.f, perbeg, perend,
method="", agedist=NULL)
```

其中, data 为患者随访资料; k 为随访时间 (年); surv.m 和 surv.f 分别为男女性 0~99 岁的条件一年生存概率, 由人群寿命表计算得到; perbeg 和 perend 是纳入分析的起始年份和截止年份; method 表示相对生存率的计算方法, 默认使用 Hakulinen 法^[17], 也可选择 Ederer II 法^[15]; agedist 命令可进行年龄调整, 可在此加上标准肿瘤患者的年龄构成^[18]来计算年龄标化生存率。

以队列法计算 1977—1981 年男性胃癌的 5 年 RSR 为例, 纳入 1972—1981 年的随访数据, 并将研究对象限定在诊断时间为 1972—1976 年的患者。语句如下:

```
cohortm<-period (subset (stomach, dy>=1972 &
dy<=1976 & sex==1), 5, probm, probf, 1972, 1981,
method="edererII")
```

完全法、现时法和实际观察到的生存率计算与之类似, 只需修改 perbeg 和 perend 两个参数, 并根据情况限制研究对象的诊断年份和性别即可。

基于现时法统计模型的实际操作过程主要分为以下 4 个步骤:

(1) 采用常规现时法计算 1~5 年的一年区间生存概率, 从中提取每个随访年份 (1~5 年) 和观察时期 (本例设定 1977—1981 年每年各为一个观察时期) 的观察死亡数 $d(\text{obs})$ 、期望死亡数 $e(\text{exp})$ 和区间内随访人数 $l(\text{py})$ 。然后计算 $d^*(\text{dstar})$ 和有效观察人数的对数 $\ln\left(l_{i,j} - \frac{d_{i,j}}{2}\right)(\text{at_risk})$, 并与随访年份 $i(y)$ 和观察时期 $j(\text{cp})$ 组成数据框“model1.df”。

(2) 自定义一个连接函数“modperiod.link”, 即用 $\ln(\mu_{i,j} - d_{i,j}^*)$ 。用“glm”函数拟合 Poisson 模型, 连接函数取刚刚自定义的“modperiod.link”, 模型中包含 2 个自变量: 随访年份 y (分类变量, $1 \leq y \leq 5$) 和观察时期 cp (连续型变量, $0 \leq \text{cp} \leq 4$), 有效观察人数的对数 (at_risk) 作为偏移项, 模型命名为“model1.fit”。

Example	Methods	Year of diagnosis	Year of follow-up															
			1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	
Example 1	Cohort analysis	1972	1	1/2	2/3	3/4	4/5	5										
		1973		1	1/2	2/3	3/4	4/5	5									
		1974			1	1/2	2/3	3/4	4/5	5								
		1975				1	1/2	2/3	3/4	4/5	5							
		1976					1	1/2	2/3	3/4	4/5	5						
		1977						1	1/2	2/3	3/4	4/5						
		1978							1	1/2	2/3	3/4						
		1979								1	1/2	2/3						
	Complete analysis	1980									1	1/2						
		1981										1						
		1972	1	1/2	2/3	3/4	4/5	5										
		1973		1	1/2	2/3	3/4	4/5	5									
		1974			1	1/2	2/3	3/4	4/5	5								
		1975				1	1/2	2/3	3/4	4/5	5							
		1976					1	1/2	2/3	3/4	4/5	5						
		1977						1	1/2	2/3	3/4	4/5						
	Period analysis	1978							1	1/2	2/3	3/4						
		1979								1	1/2	2/3						
		1980									1	1/2						
		1981										1						
		1972	1	1/2	2/3	3/4	4/5	5										
		1973		1	1/2	2/3	3/4	4/5	5									
		1974			1	1/2	2/3	3/4	4/5	5								
		1975				1	1/2	2/3	3/4	4/5	5							
	Model-based period analysis	1976					1	1/2	2/3	3/4	4/5	5						
		1977						1	1/2	2/3	3/4	4/5						
		1978							1	1/2	2/3	3/4						
		1979								1	1/2	2/3						
1980										1	1/2							
1981											1							
1972		1	1/2	2/3	3/4	4/5	5											
1973			1	1/2	2/3	3/4	4/5	5										
Later observed rates	1974			1	1/2	2/3	3/4	4/5	5									
	1975				1	1/2	2/3	3/4	4/5	5								
	1976					1	1/2	2/3	3/4	4/5	5							
	1977						1	1/2	2/3	3/4	4/5							
	1978							1	1/2	2/3	3/4							
	1979								1	1/2	2/3							
	1980									1	1/2							
	1981										1							
Example 2	Test for trend (model-based period analysis)	1973		1	1/2	2/3	3/4	4/5	5									
		1974			1	1/2	2/3	3/4	4/5	5								
		1975				1	1/2	2/3	3/4	4/5	5							
		1976					1	1/2	2/3	3/4	4/5	5						
		1977						1	1/2	2/3	3/4	4/5						
		1978							1	1/2	2/3	3/4						
		1979								1	1/2	2/3						
		1980									1	1/2						
		1981										1						
		1982											1					
		1983												1				
		1984													1			
		1985														1		
		1986															1	

The numbers within each cell represented years of follow-up since diagnosis, e.g. in 1978, a patient diagnosed in 1975 may provide survival experience of 3 or 4 years. The red frame represented the calendar period of interest in example 1, i. e. 1977–1981. The green-colored cells represented the data used by each method. Color shades were used to distinguish the different calendar periods in model-based period analysis.

图1 队列法、完全法、现时法和现时法统计模型在计算生存率和进行趋势检验时应用的生存数据

Fig 1 Survival data used by cohort method, complete method, period and model-based period analysis for survival estimation and trend test

(3) 生存率估计: 从“model1.fit”中导出随访年份和观察时期的回归系数(y_1 - y_5 和 cp), 计算最近的观察时期(1981年, $cp=4$)的1~5年一年区间生存概率, 然后累乘即得到现时法统计模型估计的最新5年RSR。

(4) 使用Delta法计算5年RSR的标准误。

计算结果如表1所示。采用队列法、完全法、现时法和现时法统计模型计算得到的1977—1981年上海市区胃癌5年RSR分别为: 男性12.77%、13.76%、14.27%和16.41%, 女性11.51%、12.04%、

12.35%和14.74%。由现时法统计模型估算的结果与后来实际观察到的男女性胃癌5年RSR(16.46%和14.23%)最为接近。实际观察值所对应的数据集

为图1的“实际观察值”部分,即在随访中可以计算的生存率。从生存率估计值的精确度来看,完全法估算结果的标准误最小。

表1 1977—1981年上海市区胃癌5年相对生存率的估计值与实际观测值的比较

Methods	Male			Female		
	5-year RSR	SE	Difference	5-year RSR	SE	Difference
Cohort analysis	12.77	0.40	-3.69	11.51	0.55	-2.72
Complete analysis	13.76	0.34	-2.7	12.04	0.46	-2.19
Period analysis	14.27	0.42	-2.19	12.35	0.56	-1.88
Model-based period analysis	16.41	0.52	-0.05	14.74	0.71	0.51
Later observed rates	16.46	0.43	-	14.23	0.57	-

RSR: Relative survival rate; SE: Standard error. Difference represented the estimated 5-year RSR of each method and the later observed rates.

实例2 实例2与实例1中现时法统计模型的计算过程基本一致,区别在于除随访年份和观察时期外,模型中还纳入了性别和年龄组这两个协变量。

第(1)步仍然是采用常规现时法计算1~5年的一年区间生存概率,从中提取观察死亡数(obs)、期望死亡数(exp)和区间内随访人数(py),并计算模型所需变量。

第(2)步是模型拟合,用“glm”函数拟合Poisson模型,连接函数取自定义的“modperiod.link”,模型中包含4个自变量:随访年份y(分类变量,1≤y≤5)、观察时期cp(0≤cp≤2,0=1978—1980年,1=1981—1983年,2=1984—1986年)、性别sex(1=男性,2=女性)以及年龄组agr(1=0~44岁,2=45~64岁,3=65~99岁)。有效观察人数(at_risk)的对数作为偏移项,模型命名为“model2.fit”。

第(3)步使用拓展包“lmtest”对回归系数cp进行Waldχ²检验。

χ²检验结果显示,χ²=19.91,P<0.001。表明在调整性别和年龄组的影响之后,胃癌的5年RSR在1978—1980年、1981—1983年和1984—1986年这3个观察时期间的趋势变化有统计学意义。

讨 论

人群肿瘤登记是肿瘤防治的一项基础性工作,通过定期收集、整理、分析和报告登记地区内恶性肿瘤的发病、现患、死亡、生存等情况,准确反映登记地区的肿瘤流行情况和变化趋势,为肿瘤的病因

学研究和防治工作提供基础数据和决策依据^[1,19]。估计肿瘤患者的生存率及进行相应的统计学检验,是全人群肿瘤登记工作的一项重要常规工作^[1,20]。由于一般人群随年龄增长有一定的“自然”死亡概率,不同地区、年代、性别、年龄组的死亡概率也不一样,为合理评价恶性肿瘤造成的超额死亡危险,方便生存率在不同人群间的比较,除了一般的观察生存率(OSR)之外,人群肿瘤登记数据的生存率计算指标通常还包括相对生存率RSR及其统计学检验^[11,20]。RSR等于观察生存率与期望生存率(expected survival rate,ESR)之比^[11]。ESR代表该地区同时期、同性别及同年龄组一般人群的“自然”生存概率,可由当地一般人群寿命表计算得到。根据期望生存率算法上的细微差别,相对生存率的估计通常有4种方法,即Ederer I^[21]、Ederer II^[15]和Hakulinen的两种方法^[17],其中Ederer II是目前最常用的方法^[22]。RSR反映了理论上肿瘤作为唯一死因时该特定人群的净生存率(net survival rate),可直接在不同人群间进行比较。

从上世纪五十年代到九十年代末,人们估计肿瘤生存率时通常仅纳入有完整随访资料的病例(队列法)。1996年Brenner等^[2]提出了现时法,作为生存率传统估计方法的有益补充,尤其是针对肿瘤长期生存资料的统计分析。常规的现时法只需稍微修改计算一年区间生存概率的时间段即可实现,目前已有基于SAS、Stata和R的宏程序或拓展包可使用^[23-25]。国内对该方法的应用主要有河南省林州市1988—2004年胃癌、食管癌的5年RSR分析^[26-28];复旦大学附属肿瘤医院2002—2006年乳腺癌5年

RSR分析^[29];以及浙江省台州市2009—2013年多种常见肿瘤的5年RSR分析^[30]。上述研究均发现现时法的生存率估计结果与之后实际的观测值最为接近。

现时法统计模型需要结合寿命表法^[12]和相对生存率^[11]的基本概念来定义和拟合广义线性模型,且标准误的计算^[13,16]也不同于常规方法,操作过程相对较为复杂。国内有学者简要介绍过现时法统计模型的基本原理^[31-32],但目前尚无应用该方法进行肿瘤生存数据分析的研究报道。现时法统计模型相比于常规现时法的最大优势在于可纳入较长年份的生存资料,在保证时效性的同时提高生存率估计的精确度。为方便与其他方法比较,本文实例1仅纳入了5年的生存资料,并以1年作为一个观察时期,每个计算区间内的有效观察例数较少,故估计的精确度不如完全法。在实际应用中,由于单独一年的生存资料容易受肿瘤登记数据延迟等情况的影响而不够稳定,通常以3或5年作为1个观察时期区间(即假定肿瘤生存率在这3或5年内稳定不变)。国际上在应用现时法统计模型时一般以10~20年的生存资料进行建模^[4-10]。本文实例2由于资料有限,仅以3年作为一个观察时期,分析了上海市胃癌5年RSR在3个观察时期(9年)间的趋势变化。Brenner等^[33]利用芬兰肿瘤登记处1978—2004年的登记资料进行比较分析,发现建模时间窗为10年时可得到与实际观测值较为接近的生存率估计值,同时标准误相比常规现时法可减少近一半;但建模时间窗延长到15年时估计效果反而变差。考虑到现时法统计模型建立在“生存率在各个观察时期之间均匀变化”这一假设前提之下,当纳入计算的时间跨度较长时,生存率的实际变化往往并不均匀。肿瘤新诊疗技术的推广应用往往会在数年后反映在患者生存率的变化上。同样,人群肿瘤筛查项目的开展可使早期检出率上升,患者的生存时间延长,从而表现为其生存率的快速提高。因此,建模时间窗并非越长越好,在实际应用中应根据肿瘤部位和当地肿瘤防控实际情况进行适当调整。

除了直接计算生存率之外,现时法统计模型的另一个优势是可以纳入其他与预后相关的协变量,如性别、年龄、肿瘤分期等,从而调整这些混杂因素对生存率的影响或进行统计学检验,以评估各个协变量对于生存率的作用。通过对模型中各个协变

量的回归系数进行假设检验来反映该协变量对RSR的作用是否有统计学意义。由于目前尚无公认的、适用于RSR的多因素分析方法,实际应用中通常只能对病例进行分层,在牺牲精确度的情况下分别在不同亚组人群中计算RSR并直接比较。现时法统计模型可以在不损失样本量的基础上调整混杂因素并进行假设检验,不失为常规分层比较方法的一种有益补充。

本文实例2结果显示,调整性别和年龄组的影响之后,胃癌的5年RSR在1978—1980年、1981—1983年和1984—1986年这3个观察时期间的趋势变化有统计学意义。与我们的实例2类似,Pulte等^[7]利用现时法统计模型分析了德国肝胆系统肿瘤在2002—2005年、2006—2009年和2010—2013年3个观察时期间的趋势变化,结果提示其5年RSR在这3个时期之间显著上升。除了对观察时期进行统计学检验以分析生存率的时间趋势之外,国外也有学者利用现时法统计模型检验性别、地区等因素对生存率的作用。例如:Eberle等^[34]利用现时法统计模型分析了2002—2010年德国肺癌人群的肿瘤登记数据,发现在调整年龄、病理学类型和TNM分期后,女性肺癌5年RSR仍显著高于男性;Majek等^[35]的研究结果显示,2002—2006年德国男女性结直肠癌患者的5年RSR分别为61.9%和64.5%,在调整年龄、分期等因素后,女性患者的预后仍然好于男性患者;Pulte等^[7]利用现时法统计模型对德国和美国的肿瘤登记资料进行分析,发现美国人群的肝细胞癌5年RSR显著高于德国人群。另外,在分析肿瘤登记生存资料时,肿瘤的年龄组划分通常采用以下标准:15~44岁、45~54岁、55~64岁、65~74岁和75岁及以上^[18]。由于本文实例2的目的在于介绍现时法统计模型如何对观察时期进行统计学检验,年龄组和性别仅作为模型中的两个协变量进行调整,因此我们只对年龄组进行了简单的3组划分。在实际应用中可根据肿瘤部位和专业知识进行相应的年龄组划分。除了将年龄组作为协变量纳入模型进行调整之外,还可进行年龄的标准化处理。PeriodR拓展包提供了简单易行的年龄标化方法,只需在第一步进行常规现时法计算时加上标准人口的年龄构成即可^[16,25]。

综上所述,在分析全人群肿瘤生存资料时,现时法统计模型可充分利用现有资料,在保证精确度

的前提下,提高生存率估计的时效性,并且能够考虑其他协变量对生存率的影响,可作为传统生存率计算方法的有益补充。但由于该方法涉及到较复杂的统计模型,且需要根据实际数据对程序中多个参数进行定义和修改,因此在进行人群肿瘤登记资料的生存分析时,应根据实际情况决定是否选用此方法并谨慎解释结果。

作者贡献声明 李卓颖 查阅文献,课题实施,编写程序,分析数据和结果解释,论文撰写和修订。蒋宇飞,谭玉婷 方法学讨论,结果解释,论文修改。项永兵 课题设计,获取资助,原始数据整理,方法学讨论,结果解释,论文修改和定稿。

利益冲突声明 所有作者均声明不存在利益冲突。

参 考 文 献

- [1] 项永兵. 肿瘤登记资料的统计分析[J]. 中国肿瘤, 2001(5): 9-11.
- [2] BRENNER H, GEFELLER O. An alternative approach to monitoring cancer patient survival[J]. *Cancer*, 1996, 78(9): 2004-2010.
- [3] BRENNER H, HAKULINEN T. Up-to-date and precise estimates of cancer patient survival: model-based period analysis[J]. *Am J Epidemiol*, 2006, 164(7): 689-696.
- [4] GONDOS A, KRILAVICIUTE A, SMAILYTE G, et al. Cancer surveillance using registry data: results and recommendations for the Lithuanian national prostate cancer early detection programme[J]. *Eur J Cancer*, 2015, 51(12): 1630-1637.
- [5] BRENNER H, GONDOS A, ARNDT V. Recent major progress in long-term cancer patient survival disclosed by modeled period analysis[J]. *J Clin Oncol*, 2007, 25(22): 3274-3280.
- [6] GONDOS A, BRAY F, HAKULINEN T, et al. Trends in cancer survival in 11 European populations from 1990 to 2009: a model-based analysis[J]. *Ann Oncol*, 2009, 20(3): 564-573.
- [7] PULTE D, WEBERPALS J, SCHRÖDER CC, et al. Survival of patients with hepatobiliary tract and duodenal cancer sites in Germany and the United States in the early 21st century[J]. *Int J Cancer*, 2018, 143(2): 324-332.
- [8] WEBERPALS J, PULTE D, JANSEN L, et al. Survival of patients with lymphoplasmacytic lymphoma and solitary plasmacytoma in Germany and the United States of America in the early 21 century[J]. *Haematologica*, 2017, 102(6): e229-e232.
- [9] YE X, MAHMUD S, SKRABEK P, et al. Long-term time trends in incidence, survival and mortality of lymphomas by subtype among adults in Manitoba, Canada: a population-based study using cancer registry data[J]. *BMJ Open*, 2017, 7(7): e015106.
- [10] WINTER A, SIRRI E, JANSEN L, et al. Comparison of prostate cancer survival in Germany and the USA: can differences be attributed to differences in stage distributions[J]. *BJU Int*, 2017, 119(4): 550-559.
- [11] 项永兵. 肿瘤流行病学研究资料的统计分析-第五讲: 相对生存率的估计[J]. 中华流行病学杂志, 1999, 20(1): 55-57.
- [12] CUTLER SJ, EDERER F. Maximum utilization of the life table method in analyzing survival[J]. *J Chronic Dis*, 1958, 8(6): 699-712.
- [13] CHIANG CL. Introduction to Stochastic Processes in Biostatistics[M]. New York: Wiley, 1968: 18.
- [14] 张薇, 鲍萍萍, 项永兵. 上海市肿瘤登记工作的历史、发展和现况[J]. 上海预防医学, 2019, 31(9): 717-723.
- [15] EDERER F, HEISE H. Instructions to IBM 650 programmers in processing survival computations[M]. MD: Bethesda, 1959.
- [16] HOLLECZEK B, BRENNER H. Model based period analysis of absolute and relative survival with R: data preparation, model fitting and derivation of survival estimates[J]. *Comput Methods Programs Biomed*, 2013, 110(2): 192-202.
- [17] HAKULINEN T. Cancer survival corrected for heterogeneity in patient withdrawal[J]. *Biometrics*, 1982, 38(4): 933-942.
- [18] CORAZZIARI I, QUINN M, CAPOCACCIA R. Standard cancer patient population for age standardising survival ratios[J]. *Eur J Cancer*, 2004, 40(15): 2307-2316.
- [19] 国家癌症中心. 中国肿瘤登记工作指导手册(2016)[M]. 北京: 人民卫生出版社, 2016: 1-2.
- [20] 项永兵, 高玉堂, 金凡. 相对生存率的统计学检验[J]. 中国卫生统计, 1996, 13(5): 15-18.
- [21] EDERER F, AXTELL LM, CUTLER SJ. The relative survival rate: a statistical methodology[J]. *Natl Cancer Inst Monogr*, 1961, 6: 101-121.
- [22] HAKULINEN T, SEPPÄ K, LAMBERT PC. Choosing the relative survival method for cancer survival estimation[J]. *Eur J Cancer*, 2011, 47(14): 2202-2210.
- [23] BRENNER H, GEFELLER O, HAKULINEN T. A computer program for period analysis of cancer patient survival[J]. *Eur J Cancer*, 2002, 38(5): 690-695.