

限制平均生存时间在临床随访研究中的应用

杨紫荆¹ 吕晶晶¹ 侯雅文² 陈征¹

¹南方医科大学公共卫生学院生物统计学系, 广州 510515; ²暨南大学经济学院统计学系, 广州 510632

通信作者: 陈征, Email: zchen@smu.edu.cn

【摘要】 在临床随访研究中, 风险比(*HR*)是衡量组间差异最常用的指标, 但由于*HR*是两组间的风险率之比, 含义较抽象, 无法对患者的生存情况做出直观的解释, 并且在使用过程中需满足比例风险假定。而此时, 限制平均生存时间(*RMST*)可以作为一种相对有效的指标或统计量, 本文介绍基于*RMST*的统计分析方法, 包括*RMST*及其差值的估计、假设检验以及回归分析, 并通过两个实例说明了*RMST*在数据分析中的应用。结果显示*RMST*可作为有效的分析工具, 其结果易于解释, 并且当比例风险假定失效时, 在组间差异的比较上相对于*HR*更有效。因此, 在疾病疗效评价和预后分析的过程中, 建议*RMST*应随*HR*同时给出, 两者相互配合、相互补充, 共同反映数据的特征。

【关键词】 限制平均生存时间; 风险比; 比例风险假定; 生存分析

基金项目: 国家自然科学基金(81673268); 广东省自然科学基金(2017A030313812, 2018A030313849); 南方医科大学大学生创新创业训练计划项目(201712121260)

DOI: 10.3760/cma.j.issn.0254-6450.2019.02.023

Application of restricted mean survival time in clinical follow-up study

Yang Zijing¹, Lyu Jingjing¹, Hou Yawen², Chen Zheng¹

¹Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou 510515, China; ²Department of Statistics, School of Economics, Jinan University, Guangzhou 510632, China

Corresponding author: Chen Zheng, Email: zchen@smu.edu.cn

【Abstract】 In clinical follow-up studies, hazard ratio (*HR*) is routinely used to quantify the differences between-groups, however, it is being estimated by the Cox procedure. *HR*, the ratio of two hazard functions has abstract meaning only and is in lack of the context to give an intuitive explanation of the survival of patients and the assumption of proportional hazards (*PH*) must be satisfied. Under this context, the restricted mean survival time (*RMST*) can be used as a relatively effective measure or index of statistics. This paper introduces the *RMST*-based statistical analysis methods, including estimation of *RMST* and its difference, hypothesis testing and regression analysis. The application of *RMST* in data analysis is also introduced. All the evidence demonstrates that *RMST* can be used as an effective analytical tool with straightforward interpretation. *RMST* is also more effective than *HR* in comparing differences between groups, when non-*PH* is observed. Therefore, *RMST* is suggested to be stated along with *HR* in the process of disease efficacy evaluation and prognosis analysis. Cooperation and complement of the two, a precise reflection on the characteristics of data can be expected.

【Key words】 Restricted mean survival time; Hazard ratio; Proportional hazards assumption; Survival analysis

Fund programs: National Natural Science Foundation of China (81673268); Natural Science Foundation of Guangdong Province (2017A030313812, 2018A030313849); College Students' Innovation and Entrepreneurship Training Program of Southern Medical University (201712121260)

DOI: 10.3760/cma.j.issn.0254-6450.2019.02.023

在临床随访研究中, 风险比(hazard ratio, *HR*)是衡量组间差异最常用的指标。但由于*HR*是两组间的风险率之比, 其含义较抽象, 因此无法对患者的

生存情况做出直观的解释^[1-3]。同时, 在*HR*的使用过程中需满足比例风险(proportional hazards, *PH*)假定, 即*HR*不随时间发生变化, 而在长时间的随访过

程中常常出现比例风险假定失效的问题,此时继续使用HR作为评价指标将不再合理^[4-5]。因此,有研究者建议使用其他统计量或指标来避免比例风险假定的前提条件,如中位生存时间和平均生存时间等^[6]。然而,当删失率较高(如>50%)时,中位生存时间无法得到;同时,由于随访中删失的存在和随访时间的限制,导致无法获得t趋于无穷大时的随访数据,因此严格的平均生存时间也难以获得。为解决上述问题,Irwin^[7]提出将平均生存时间的积分上限由无穷大限制到某一固定的时间点t*,定义为限制平均生存时间(restricted mean survival time, RMST),即到某一特定时间点t*患者的平均生存时间,或t*年的预期寿命^[8]。例如在随访研究中可得出如下类似结论,“从入组时刻起,患者1年内的平均生存时间为9个月”或是“与传统疗法相比,接受新型治疗的患者五年预期寿命将提高3个月”。本文将介绍RMST的估计以及相应的统计推断方法,并结合实例进行分析说明。

基本原理

1. RMST的定义:假设T为生存时间,对于某个限制时间点t*(t*>0),使得Pr(T≥t*)>0,限制生存时间(restricted survival time)为X=min(T,t*),X的期望即为RMST,通常用μ(t*)来表示:

$$\mu(t^*) = E(X) = E[\min(T, t^*)] = \int_0^{t^*} S(t) dt \quad (1)$$

可以看出,μ(t*)在数值上等价于t=0到t=t*时生存曲线S(t)下的面积,当t*→∞时,μ(t*)即为平均生存时间。X对应的方差为

$$\text{Var}(X) = 2 \int_0^{t^*} tS(t) dt - \left[\int_0^{t^*} S(t) dt \right]^2 \quad (2)$$

2. RMST的估计:通常μ(t*)可根据Kaplan-Meier法进行计算

$$\hat{\mu}(t^*) = \int_0^{t^*} \hat{S}_{KM}(t) dt \quad (3)$$

其中 $\hat{S}_{KM}(t)$ 是Kaplan-Meier估计下的生存率。设 $t_{(1)} < t_{(2)} < \dots < t_{(D)}$ 是D个观察到的发生兴趣事件的时间, $d_{(i)}$ 是 $t_{(i)}$ 时刻发生兴趣事件的人数, $n_{(i)}$ 是 $t_{(i)}$ 时刻的风险集人数,可以得到近似方差为:

$$\text{Var}(\hat{\mu}(t^*)) = \sum_{i=1}^D \left[\int_{t_{(i)}}^{t^*} \hat{S}_{KM}(t) dt \right]^2 \frac{d_{(i)}}{n_{(i)}(n_{(i)} - d_{(i)})} \quad (4)$$

对应的标准误为:

$$SE(\hat{\mu}(t^*)) = \sqrt{\text{Var}(\hat{\mu}(t^*))} \quad (5)$$

RMST的双侧1-α置信区间估计为:

$$\hat{\mu}(t^*) - Z_{1-\alpha/2} SE(\hat{\mu}(t^*)), \hat{\mu}(t^*) + Z_{1-\alpha/2} SE(\hat{\mu}(t^*)) \quad (6)$$

其中 $Z_{1-\alpha/2}$ 是标准正态分布的100(1-α/2)百分位数。

3. 假设检验:假设有两组数据,如试验组和对照组,根据公式(3)和(5)可分别计算每组RMST的估计值及其标准误 $\hat{\mu}_0, SE(\hat{\mu}_0)$ 和 $\hat{\mu}_1, SE(\hat{\mu}_1)$ 则组间差值 $\hat{\Delta} = \hat{\mu}_1 - \hat{\mu}_0$,对应的标准误为 $SE(\hat{\Delta}) = \sqrt{SE(\hat{\mu}_0)^2 + SE(\hat{\mu}_1)^2}$ 。相应的假设检验为: $H_0: \Delta = \mu_1 - \mu_0 = 0, H_1: \Delta \neq 0$ 。当 H_0 成立时,在大样本情况下,检验统计量 $Z = \hat{\Delta} / SE(\hat{\Delta})$ 近似服从标准正态分布。组间差值的双侧1-α置信区间估计为:

$$\hat{\Delta} - Z_{1-\alpha/2} SE(\hat{\Delta}), \hat{\Delta} + Z_{1-\alpha/2} SE(\hat{\Delta}) \quad (7)$$

其中 $Z_{1-\alpha/2}$ 是标准正态分布的100(1-α/2)百分位数。

4. 回归模型:在疾病的预后分析中,可以使用“伪观测”(或“伪值”)来评估协变量对RMST的影响^[9],也就是将RMST的伪值作为广义线性模型(GLMs)的因变量对协变量的效应进行建模^[10]。伪值的定义:设生存时间 $T_1, T_2, T_3, \dots, T_n$ 独立同分布, θ 是兴趣参数RMST,根据全部样本可以计算得到 θ 的无偏(或近似无偏)估计 $\hat{\theta} = \hat{\mu}(t^*) = \int_0^{t^*} \hat{S}(t) dt$ 。在舍弃第i个观测值之后,根据剩余的(n-1)个样本用同样的方式可以计算出 θ 的一个估计值 $\hat{\theta}^{-i}$,称之为 θ 基于 $T_j(j \neq i)$ 的“留一法(leave-one-out)”估计,则第i个伪值可定义为:

$$\hat{\theta}_i = n \cdot \hat{\theta} - (n-1) \hat{\theta}^{-i} \quad (8)$$

假设存在协变量 Z_1, Z_2, \dots, Z_m ,在计算出伪值后,对于给定的协变量 Z_i 可相应建立起 θ 的回归模型进行估计和预测:

$$g(\theta_i) = \beta^T Z_i \quad (9)$$

其中 $g(\cdot)$ 是一个连接函数,本文中所使用的是恒等函数(identity link function),同时加入截距 β_0 。可通过标准的广义估计方程求解系数 β :

$$U(\beta) = \sum_i U_i(\beta) = \sum_i \left(\frac{\partial}{\partial \beta} g^{-1}(\beta^T Z_i) \right)^T V_i^{-1} \left(\hat{\theta}_i - g^{-1}(\beta^T Z_i) \right) = 0 \quad (10)$$

V_i 是 $\hat{\theta}_i$ 的方差,并根据Sandwich法计算出 β 的方差。

实例分析

例1. 单因素分析。在一项评价肾功能不全患者初次发生导管口感染时间的研究中^[11],共收集119名患者的数据,其中43人进行外科放置导管,76人进

行皮下放置导管。终点事件为患者初次发生通道口感染,导尿管失效或试验结束时终点事件尚未发生为右删失,2组的删失率分别为 65.1%和 85.5%。Cox 模型得出 $HR=0.542(95\%CI:0.249 \sim 1.182, P=0.124)$,说明这两种治疗方法的疗效差异无统计学意义,经 Grambsch-Therneau 检验发现两组风险率不满足比例风险假定 ($\chi^2=8.700, P=0.003$),同时 2 条生存曲线在第 8 个月左右相交(图 1),而在交叉点之后两组可能会存在疗效上的差异,因此 Cox 模型的结果可能没有反映真实情况。

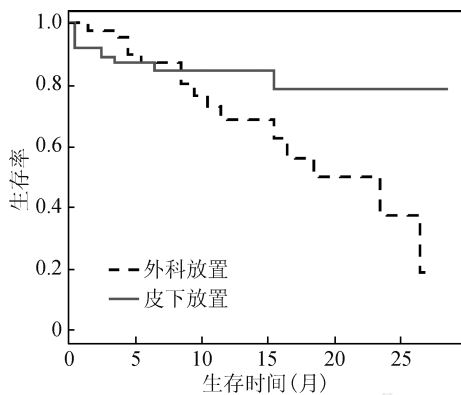


图 1 肾透析患者进行外科和皮下放置导管的生存曲线

在 RMST 的分析中,选择 2 个不同的 t^* ,其中第 8 个月为 2 条生存曲线的近似交叉点所对应的时间,27.5 个月时则为两组最大观察时间(the last observed time)的最小值,这里的观察时间指的是发生终点事件或右删失的时间。表 1 给出了不同 t^* 的分析结果,可以看出,在治疗短期 $t^*=8$ 个月时,外科放置导管的疗效略优于皮下放置导管,但检验结果显示两组的短期疗效差异无统计学意义 ($Z=1.080, P=0.280$);而

对于治疗的整体效果,在给定 $t^*=27.5$ 个月时,可以得到进行皮下放置导管的患者在未来 27.5 个月时的平均生存时间约为 22.9 个月时,而进行外科放置导管的患者平均生存时间约为 18.5 个月时,即进行皮下放置导管相比于外科放置导管将患者未来 27.5 个月时的平均生存时间提高了约 4.3 个月时,组间差异有统计学意义 ($Z=2.048, P=0.041$)。

例 2. 多因素分析。数据来源于北部中心肿瘤治疗协助组(North Central Cancer Treatment Group, NCCTG)提供的关于晚期肺癌患者生存的数据^[12],共包含 228 名患者的信息,删除存在未知变量的数据,最终有 227 名患者纳入分析,终点事件为患者死亡,其余为右删失。将性别(男性赋值为 1,女性为 0)、年龄(<65 岁为 1, ≥ 65 岁为 0)以及 Karnofsky 评分(<80 分为 1, ≥ 80 分为 0)作为自变量,应用伪值回归模型分析该数据,以探讨晚期肺癌患者预后的影响因素。

分析发现,若将 t^* 设置为最长随访时间 2.8 年,不同性别对 RMST 的影响差异有统计学意义 ($Z=3.035, P=0.002$),男性晚期肺癌患者在未来 2.8 年里平均比女性患者少活约 0.34 年,而患者的年龄对其预后的影响差异无统计学意义 ($Z=1.444, P=0.149$)(表 2)。Karnofsky 评分对其预后的影响差异有统计学意义 ($Z=2.090, P=0.037$),评分 ≥ 80 分患者(即能够进行正常活动)的 RMST 相比于其余患者将延长约 0.26 年,尽管 HR 也能够得出类似的结论,但由于检验发现比例风险假定失效 ($\chi^2=7.060, P=0.008$),此时该变量对患者预后的影响可能随时间发生变化,HR 并不是一个恒定的值,在实际解释中

表 1 基于 HR 和 RMST 对肾透析患者疗效的单因素分析

模型	t^*	皮下放置(95%CI)	外科放置(95%CI)	HR/RMSTd(95%CI)	统计量	P 值
HR(Cox 模型)	-	-	-	0.542*(0.249 ~ 1.182)	1.539	0.124
RMST	8	7.108(6.582 ~ 7.634)	7.486(7.046 ~ 7.925)	-0.378*(-1.063 ~ 0.308)	1.080	0.280
	27.5	22.865(20.281 ~ 25.448)	18.527(15.275 ~ 21.778)	4.338*(0.185 ~ 8.491)	2.048	0.041

注: *HR 的分析结果; ^bRMSTd 为两组 RMST 的差值

表 2 基于 Cox 回归模型和 RMST 回归模型对晚期肺癌患者预后的多因素分析

变量(样本量)	Cox 比例风险模型			RMST 回归模型			
	HR 值(95%CI)	Z 值	P 值	P 值 ^a	β 值(95%CI)	Z 值	P 值
截距					1.221(1.004 ~ 1.438)	11.033	<0.001
性别							
男(137)	1.670(1.203 ~ 2.318)	3.061	0.002	0.119	-0.338(-0.556 ~ -0.120)	3.035	0.002
女(90)	1.000						
年龄组(岁)							
<65(127)	0.815(0.590 ~ 1.124)	1.248	0.212	0.663	0.148(-0.053 ~ 0.349)	1.444	0.149
≥ 65 (100)	1.000						
Karnofsky 评分							
<80(57)	1.439(1.013 ~ 2.044)	2.033	0.042	0.008	-0.261(-0.505 ~ -0.016)	2.090	0.037
≥ 80 (170)	1.000						

注: ^aGrambsch-Therneau 风险率成比例假设检验的 P 值

相对复杂。

讨 论

本文介绍了 *RMST* 的定义、估计、假设检验和回归分析方法,并将该指标与 *HR* 进行了比较。实例 1 和 2 都表明,相较于 *HR*, *RMST* 在临床上更直观和易解释。这是因为 *RMST* 差值衡量了某 t^* 时处理因素对生存时间的影响,是一个相对的效应度量,同时每组 *RMST* 的值又是对生存时间的绝对度量。而 *HR* 只是一个相对指标,既不反映生存时间也不反映生存率,即使不同处理对生存时间的绝对影响很小, *HR* 也可以非常大,因此它需要伴随其他统计量/指标共同给出,如中位生存时间、特定时间点的生存率等。在很多情况下, *RMST* 的差值可能看起来“很小”,但它比 *HR* 更具现实性和临床意义。

在实例 1 中,基于 *RMST* 检验法在不满足比例风险假定的情况下比 Cox 模型更易发现组间差异,这是因为 *HR* 作为评价指标时,只有在满足比例风险假定的情况下对应的统计方法才有更好的性能,否则会降低、甚至产生错误的结论^[13]。而 *RMST* 不需要任何前提假定,通过直接对生存函数进行积分求得,在数值上等价于从 0 到 t^* 时生存曲线 $S(t)$ 下的面积,考虑的是 t^* 前的整个生存分布而非单一时间点的生存率,相比于中位生存时间等指标来说利用了更多的信息。同时,基于 *RMST* 的检验法除了使用两组 *RMST* 的差值外,还可以对相应的比值进行检验,如设置 *RMST* 的比值为试验组与对照组的 *RMST* 之比,该值 >1 说明试验组疗效优于对照组,反之则认为对照组优于试验组。实例 1 中当给定 t^* 为 27.5 个月时,计算出外科放置与皮下放置这两组 *RMST* 的比值以及相应的可信区间为 1.234 (95% *CI*: 1.002 ~ 1.521, $P=0.048$),在临床上可解释为:平均来说,进行外科放置导管的患者在未来 27.5 个月时的生存时间比进行皮下放置导管的患者多 23.4%。

但在使用 *RMST* 的过程中也需要注意下列问题。 t^* 应在试验设计阶段给定,并合理选择具有临床意义且接近研究随访的结束时间,这样可以使试验更加完整且包含大多数的生存结局。这也是 *RMST* 的优势之一, *RMST* 将时间的维度纳入到分析结果中^[13],根据研究目的可以选择不同的 t^* 来检验各时期的处理效果。实例 1 中,当使用单一的 *HR* 无法反映出处理效应随时间的变化时,选择一个较小的 t^* 可以分别得到处理组在短期和长期的疗效差异,如给定 t^* 为 8 个月时,除了给出前 8 个月的 *RMST* 外,还

可以计算出 8 个月后两组的 *RMST* 之差为 4.716 个月,从而区分该时刻前期和后期的治疗差异。但是,当风险集人数较少时, Kaplan-Meier 生存估计会不稳定^[8],通过直接对 Kaplan-Meier 生存曲线进行积分计算得出的 *RMST* 可能会存在误差,此时可以考虑采用更为稳健的估计方法例如灵活参数模型^[14]等。

综上所述,虽然基于 *HR* 的统计分析方法(如 log-rank 检验和 Cox 比例风险回归模型等)在比例风险假定不成立的情况下检验效能会有所下降,且在临床解释上存在一定的问题,但在成比例的情况下检验效能仍为最佳^[2,15]。因此,在疾病疗效评价和预后分析的过程中,建议 *RMST* 随 *HR* 同时给出,两者相互配合、补充,共同反映数据的特征。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

- [1] Horiguchi M, Pak K, Mikami M, et al. Issues of the hazard ratio estimate and application of the restricted mean survival time to a non-inferiority study (New Advances in Statistical Inference and Its Related Topics) [J]. RIMS Kokyuroku, 2015, 1954: 1-14.
- [2] Huang B, Kuan PF. Comparison of the restricted mean survival time with the hazard ratio in superiority trials with a time-to-event end point [J]. Pharm Stat, 2017, 17(3): 202-213. DOI: 10.1002/pst.1846.
- [3] Trinquart L, Jacot J, Conner SC, et al. Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials [J]. J Clin Oncol, 2016, 34(15): 1813-1819. DOI: 10.1200/jco.2015.64.2488.
- [4] Uno H, Wittes J, Fu HD, et al. Alternatives to hazard ratios for comparing the efficacy or safety of therapies in noninferiority studies [J]. Ann Intern Med, 2015, 163(2): 127-134. DOI: 10.7326/M14-1741.
- [5] Uno H, Claggett B, Tian L, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis [J]. J Clin Oncol, 2014, 32(22): 2380-2385. DOI: 10.1200/jco.2014.55.2208.
- [6] Saad ED, Zalcberg JR, Pécron J, et al. Understanding and communicating measures of treatment effect on survival: can we do better? [J]. J Natl Cancer Inst, 2018, 110(3): 232-240. DOI: 10.1093/jnci/djx179.
- [7] Irwin JO. The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice [J]. J Hyg (Lond), 1949, 47(2): 188-189. DOI: 10.1017/S0022172400014443.
- [8] Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt [J]. Stat Med, 2011, 30(19): 2409-2421. DOI: 10.1002/sim.4274.
- [9] Andersen PK, Hansen MG, Klein JP. Regression analysis of restricted mean survival time based on pseudo-observations [J]. Lifetime Data Anal, 2004, 10(4): 335-350. DOI: 10.1007/s10985-004-4771-0.
- [10] Klein JP, Gerster M, Andersen PK, et al. SAS and R functions to compute pseudo-values for censored data regression [J]. Comput Methods Programs Biomed, 2008, 89(3): 289-300. DOI: 10.1016/j.cmpb.2007.11.017.
- [11] Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data [M]. 2nd ed. New York: Springer, 2003.
- [12] Loprinzi CL, Laurie JA, Wieand HS, et al. Prospective evaluation of prognostic variables from patient-completed questionnaires. North Central Cancer Treatment Group [J]. J Clin Oncol, 1994, 12(3): 601-607. DOI: 10.1200/jco.1994.12.3.601.
- [13] Royston P, Parmar MKB. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome [J]. BMC Med Res Methodol, 2013, 13: 152. DOI: 10.1186/1471-2288-13-152.
- [14] Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects [J]. Stat Med, 2002, 21(15): 2175-2197. DOI: 10.1002/sim.1203.
- [15] Horiguchi M, Cronin AM, Takeuchi M, et al. A flexible and coherent test/estimation procedure based on restricted mean survival times for censored time-to-event data in randomized clinical trials [J]. Stat Med, 2017, 37(15): 2307-2320. DOI: 10.1002/sim.7661.

(收稿日期: 2018-07-24)
(本文编辑: 李银鸽)