

# 固定点处组间生存率比较的统计检验法

陈金宝 邱李斌 王北琪 曾洁 陈征

**【导读】** 组间生存曲线的比较是研究疾病预后的重要内容,目前常用Log-rank检验,但当生存曲线交叉或者两曲线后期开口较大,即比例假设不成立时,采用该法则失效。为此本文讨论适用这种情形的比较法。介绍两条生存曲线某固定点处生存率比较的5种(经典法、对数转换法、cloglog转换法、反正弦转换法和logit转换法)统计检验方法。通过实例分析发现,若用组间生存曲线整体检验的Log-rank检验法或Two-stage检验法均很难对两组生存率比较给出合理的结果,而选取几个固定点进行检验时,可得到固定时间点上组间生存率差异是否有统计学意义。在5种固定点时刻的检验方法中,以cloglog转换法最佳。

**【关键词】** 生存曲线; 组间生存率比较; Kaplan-Meier估计; 固定点

**Statistical methods for comparing survival rates at a fixed time point** Chen Jinbao, Qiu Libin, Wang Beiqi, Zeng Jie, Chen Zheng. The Department of Biostatistics, School of Public Health and Tropical Medicine, Southern Medical University, Guangzhou 510515, China

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

This work was supported by grants from the National Natural Science Foundation of China (No. 81202288), the Science and Technology Plan of Guangzhou (No. 2012J5100023), the Special Research Fund for the Doctoral Program of Higher Education (No. 20114433120010), the College Students' Innovative Entrepreneurial and Training Plan of Guangdong Province (No. 1212113039).

**【Introduction】** Comparison of survival curves between two groups is an important part of disease prognosis study. Log-rank test is commonly used, but when the two curves' later intersecting opening is too large, the proportion of assumptions is not established, thus the Log-rank test is ineffective. We introduces five statistical tests to compare two survival curves at a fixed time points: classic method, logarithmic transformation, cloglog transformation, arcsine transformation and logit transformation. Through the study we found that if the overall survival curves are difficult to compare between groups tested with Log-rank test method or Two-stage test method, the fixed time point test can effectively determine whether there was significant difference in survival rate at a fixed time point. Among the five fixed time point tests, cloglog transformation could give more precise result.

**【Key words】** Survival curves; Comparison of survival rates between groups; Kaplan-Meier estimate; Fixed time point

疾病预后分析是临床流行病学的重要研究内容,其中估计生存率常用的方法是寿命表或是Kaplan-Meier估计,如胃癌患者治疗后5年( $t=5$ )的生存率为25% [ $S(t)=0.25$ ]。同样,不同暴露环境/处理方式下两样本生存率的比较也是重要的研究问题,通常采用Log-rank检验推断两生存曲线整体间(整个观察期间)是否有差异,如放疗组和放化疗联合组在预后5年生存率有无差异,若生存曲线出现

交叉情况,由于不满足成比例假设,则Log-rank检验不适用<sup>[1-2]</sup>。例如,在一项胃肠道肿瘤治疗法的8年随访研究中<sup>[1]</sup>,对局部无法切除的胃癌患者分别进行单纯化疗和联合放化疗,由于两条生存曲线大约在第3年时交叉(图1),并经Grambsch-Therneau检验<sup>[2]</sup>,两组累积危险率不成比例( $\chi^2=13.10, P<0.001$ ),因此Log-rank检验结果( $\chi^2=0.23, P=0.630$ )不可靠。在出现生存曲线交叉的情况下,李慧敏等<sup>[3]</sup>建议采用Two-stage检验可得到两组间生存率的差异有统计学意义( $P=0.025$ )的结果,结合图1,依旧很难给出在整体上哪种治疗法生存率更高的结论,并且在临床研究中更加感兴趣是第3、5年等某些特别/固定时间点上两种治疗法疗效的预后差异,如放疗和化疗在第1、3、5年的生存率是否有差异。

DOI: 10.3760/cma.j.issn.0254-6450.2015.02.019

基金项目:国家自然科学基金(81202288);广州市科技计划(2012J5100023);高等学校博士学科点专项科研基金(20114433120010);广东省大学创业创新训练计划(1212113039)

作者单位:510515 广州,南方医科大学公共卫生与热带医学学院生物统计系

通信作者:陈征, Email: zchen@smu.edu.cn

Weeda 等<sup>[4]</sup>在生存曲线交叉时,使用 Z 检验比较第 3 年时的组间差异,但针对带右删失的生存数据采用 Z 检验并不适合。针对上述存在的问题,本文介绍 5 种在固定点比较生存率的统计推断方法,分别为经典法、对数转换法、cloglog 转换法、反正弦法和 logit 转换法,并结合实例分析说明。

### 基本原理

这里主要考虑两条生存曲线在某个固定时间点的比较。假设第 k 组有样本量 n<sub>k</sub>,其中 k=1,2, t<sub>ki</sub> 表示第 k 组中的第 i 个事件发生的时间,且 t<sub>k1</sub><t<sub>k2</sub><...<t<sub>kn</sub>,d<sub>ki</sub> 表示第 k 组中在时间 t<sub>ki</sub> 发生事件数, Y<sub>ki</sub> 表示第 k 组在时间点 t<sub>ki</sub> 风险人数。Kaplan-Meier 估计及其方差分别表示为  $\hat{S}_k(t) = \prod_{t_{ki} \leq t} (1 - d_{ki}/Y_{ki})$  和  $\hat{V}[\hat{S}_k(t)] = \hat{S}_k(t)^2 \hat{\sigma}_k(t)^2$ , 其中  $\hat{\sigma}_k(t)^2 = \sum_{t_{ki} \leq t} d_{ki}/Y_{ki}(Y_{ki} - d_{ki})$ , 累积危险率及其方差估计值分别是  $\hat{H}_k(t) = \sum_{t_{ki} \leq t} d_{ki}/Y_{ki}$  和  $\hat{V}[\hat{H}_k(t)] = \sum_{t_{ki} \leq t} d_{ki}/Y_{ki}^2$ 。

比较在时间点 t 的两组生存率,原假设 H<sub>0</sub> 是在某固定时间点 t, 两组生存率相等,即 S<sub>1</sub>(t)=S<sub>2</sub>(t)。以下提供 5 种检验法<sup>[5]</sup>。

#### 1. 经典法:

$$\chi^2_1 = \frac{[\hat{S}_1(t) - \hat{S}_2(t)]^2}{\hat{S}_1(t)^2 \hat{\sigma}_1(t)^2 + \hat{S}_2(t)^2 \hat{\sigma}_2(t)^2}$$

在原假设成立下,此检验的统计量近似服从自由度为 1 的  $\chi^2$  分布。类似的可基于危险率函数 Nelson-Aalen 估计,构建检验统计量为

$$\chi^2_1 = \frac{[\hat{H}_1(t) - \hat{H}_2(t)]^2}{\hat{V}[\hat{H}_1(t)] + \hat{V}[\hat{H}_2(t)]}$$

检验统计量是基于  $\hat{S}_k(t)$  或  $\hat{H}_k(t)$  的某个恰当转换(或说函数)  $\phi$ , 因此通过 delta 法, 可得  $\phi[\hat{S}_k(t)]$  近似服从均值为  $\phi[S_k(t)]$ , 方差为  $\hat{V}[\hat{S}_k(t)] \{ \phi'[\hat{S}_k(t)] \}^2$  的正态分布。通过危险率 Nelson-Aalen 估计的转换也可以得到类似结果。对于生存函数的 Kaplan-Meier 估计来说,基于方差的稳健转换被用来更好的呈现转换过后的生存函数的置信区间<sup>[6]</sup>。在此,可以认为在原假设成立下,以下一系列的转换得到的统计量都近似服从  $\chi^2$  分布。

2. 对数转换:基于 Nelson-Aalen 估计值,对生存函数进行对数转换后得到检验统计量

$$\chi^2_2 = \frac{\{ \log[\hat{S}_1(t)] - \log[\hat{S}_2(t)] \}^2}{\hat{\sigma}_1(t)^2 + \hat{\sigma}_2(t)^2}$$

3. cloglog 转换:即基于生存函数的  $\log[-\log(\cdot)]$

转换,通常用于构建生存函数的置信区间<sup>[6]</sup>和置信带<sup>[7]</sup>,在小样本量(如 n=25),且删失率高(如 50%)时,此转换法也可给出准确覆盖率作为 95%CI。在原假设成立的条件下,检验统计量为

$$\chi^2_3 = \frac{\{ \log(-\log[\hat{S}_1(t)]) - \log(-\log[\hat{S}_2(t)]) \}^2}{\hat{\sigma}_1(t)^2 / \{ \log[\hat{S}_1(t)] \}^2 + \hat{\sigma}_2(t)^2 / \{ \log[\hat{S}_2(t)] \}^2}$$

4. 反正弦转换:此统计量的构造是基于反正弦根(arc sine-square root)的转换法,同 cloglog 转换法也可给出小样本 95% 的覆盖率作为其 CI<sup>[6]</sup>。统计量为

$$\chi^2_4 = \frac{\{ \arcsin \sqrt{\hat{S}_1(t)} - \arcsin \sqrt{\hat{S}_2(t)} \}^2}{\hat{v}_1(t) + \hat{v}_2(t)}$$

其中,  $\hat{v}_k(t) = \frac{\hat{S}_k(t) \hat{\sigma}_k(t)^2}{4[1 - \hat{S}_k(t)]}$ , k=1,2

5. logit 转换:生存率一般不服从正态分布,logit 转换有助于调整数据的形态,且可以较好地处理非线性相关的指标与增长性的分析。统计量为

$$\chi^2_5 = \frac{\left\{ \log \frac{\hat{S}_1(t)}{1 - \hat{S}_1(t)} - \log \frac{\hat{S}_2(t)}{1 - \hat{S}_2(t)} \right\}^2}{\frac{\hat{\sigma}_1(t)^2}{[1 - \hat{S}_1(t)]^2} + \frac{\hat{\sigma}_2(t)^2}{[1 - \hat{S}_2(t)]^2}}$$

### 实例分析

本文提供 2 个实例(其中例 1 表现为生存曲线交叉,例 2 为生存曲线前期基本重合,后期分开)比较上述 5 种固定点生存率统计方法间的差异。

**【例 1】** 一项胃肠道肿瘤治疗法的随访研究中共计 90 例胃癌患者,分别进行单纯化疗和联合放化疗,每组样本量为 45 例(右删失率分别为 4.4% 和 13.3%),终点事件为死于胃癌,到研究截止时间为止未发生终点事件的胃癌患者为右删失(图 1)。两组的中位生存时间分别为 499 d 和 254 d。表 1 中 5 种检验方法对两条生存曲线几个时间点处的比较结果,均显示两组间第 1 年生存率的差异有统计学意义,在治疗胃肠道肿瘤上单纯化疗的生存率较联合放化疗的生存率高,而第 3、5 年生存率的差异无统计学意义。若用 Log-rank 检验或 Two-stage 检验均很难给出合理的结果。

**【例 2】** 一项预防腺性息肉瘤的研究<sup>[8]</sup>,对曾患结肠直肠癌腺瘤的患者分别服用万络和安慰剂,终点事件是发生心血管事件(如心脏病及中风),到研究截止时尚未发生终点事件的结肠直肠癌腺瘤患者为右删失,结果显示入组时两组患者出现心血管事件的概率基本一致,但 18 个月后服用万络的患者出现心血管事件的概率高于服安慰剂者。由于原文数据无法获得,因此参照 Royston 和 Mahesh<sup>[9]</sup>的方法模拟

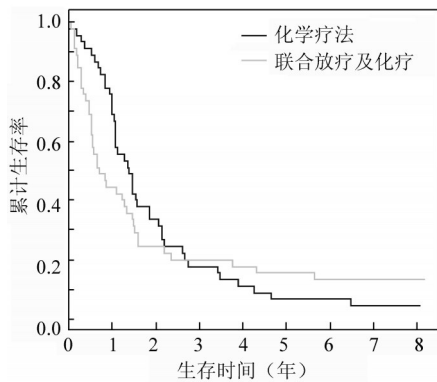


图1 胃癌患者分别进行单纯化疗和联合放疗及化疗的生存曲线

表1 不同检验方法在[例1]的应用结果

方法	第1年	第3年	第5年
经典法 <sup>a</sup>	0.016(5.83)	0.788(0.07)	0.175(1.84)
对数转换法 <sup>a</sup>	0.024(5.08)	0.788(0.07)	0.197(1.66)
cloglog转换法 <sup>a</sup>	0.022(5.28)	0.788(0.07)	0.177(1.82)
反正弦转换法 <sup>a</sup>	0.018(5.60)	0.788(0.07)	0.171(1.87)
logit转换法 <sup>a</sup>	0.021(5.35)	0.788(0.07)	0.192(1.71)
Log-rank检验 <sup>b</sup>	0.630(0.23)		
Two-stage检验 <sup>b</sup>	0.025		

注:括号内为P值对应的 $\chi^2$ 统计量,检验水准 $\alpha = 0.05$ ; <sup>a</sup>某固定时间点处的组间生存率比较; <sup>b</sup>生存曲线整体间的比较

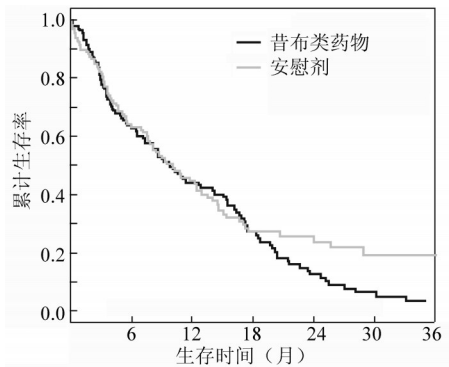


图2 结肠直肠腺瘤患者分别服用昔布类药物和安慰剂的生存曲线

表2 不同检验方法在[例2]的应用结果

方法	第12个月	第18个月	第24个月	第30个月
经典法 <sup>a</sup>	0.930(0.01)	0.998(0.00)	0.048(3.91)	0.019(5.54)
对数转换法 <sup>a</sup>	0.930(0.01)	0.998(0.00)	0.054(3.72)	0.022(5.23)
cloglog转换法 <sup>a</sup>	0.930(0.01)	0.998(0.00)	0.048(3.90)	0.016(5.79)
反正弦转换法 <sup>a</sup>	0.930(0.01)	0.998(0.00)	0.047(3.96)	0.015(5.97)
logit转换法 <sup>a</sup>	0.930(0.01)	0.998(0.00)	0.051(3.80)	0.020(5.40)
Log-rank检验 <sup>b</sup>	0.223(1.48)			
Two-stage检验 <sup>b</sup>	0.035			

注:同表1

的检验法时,或者仅需要比较某些固定点处组间生存率差异时,可使用本文以上介绍的固定点时刻生存率比较检验法。其中,如 Klein 等<sup>[5]</sup>通过 Monte Carlo 模拟研究表明,结合第一类和第二类错误的结果,cloglog 转换的检验性能最好,其次是反正弦转换、对数转换和 logit 转换,经典法相对较差,与采用这 5 种转换方法计算生存率的 CI 值时分析结果类似<sup>[6]</sup>,建议使用 cloglog 转换法。

参 考 文 献

- [1] Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. 2nd ed[M]. New York (NY): Springer, 2003.
- [2] Grambsch P, Therneau T. Proportional hazards tests and diagnostics based on weighted residuals [J]. Biometrika, 1994, 81: 515-526.
- [3] Li HM, Han D, Chen Z, et al. Comparison and choice of statistical inference methods for two crossing survival curves [J]. Chin J Health Stat, 2013, 30(5): 668-672. (in Chinese) 李慧敏, 韩栋, 陈征, 等. 生存曲线交叉时统计推断的比较和选择 [J]. 中国卫生统计, 2013, 30(5): 668-672.
- [4] Weeda VB, Murawski M, McCabe AJ, et al. Fibrolamellar variant of hepato-cellular carcinoma does not have a better survival than conventional hepato-cellular carcinoma-Results and treatment recommendations from the Childhood Liver Tumour Strategy Group (SIOPEL) experience [J]. Eur J Cancer, 2013, 49(12): 2698-2704.
- [5] Klein JP, Logan B, Harhoff M, et al. Analyzing survival curves at a fixed point in time [J]. Stat Med, 2007, 26: 4505-4519.
- [6] Xiang YB, Gao YT, Jin F, et al. Five methods for the estimation of confidence intervals on survival rate [J]. Chin J Epidemiol, 1995, 16(5): 306-309. (in Chinese) 项永兵, 高玉堂, 金凡, 等. 生存率置信区间的五种估计方法 [J]. 中华流行病学杂志, 1995, 16(5): 306-309.
- [7] Borgon φ, Liestøl K. A note on confidence intervals and bands for the survival curve based on transformations [J]. Scand J Stat, 1990, 17: 35-41.
- [8] Lagakos SW. Time-to event analyses for long-term treatments-the APPROVe trail [J]. N Engl J Med, 2006, 355: 113-117.
- [9] Royston P, Mahesh KBP. 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.

(收稿日期: 2014-09-27)  
(本文编辑: 张林东)

了 Lagakos<sup>[8]</sup>文中的临床数据(图2)。万络组和安慰剂组的样本量分别为 132 人和 130 人, 对应右删失率为 22.0% 和 37.7%, 中位生存时间为 9.71 个月和 10.04 个月。Log-rank 检验结果为两组间整体的差异无统计学意义 ( $\chi^2=1.48, P=0.223$ ), 但由于 Grambsch-Therneau 检验显示两组不满足比例假设 ( $\chi^2=4.45, P=0.035$ )<sup>[2]</sup>, 图2同时显示两组短期(约 18 个月前)和长期(约 18 个月后)效应似乎不同, 因此 Log-rank 检验结果不可靠。采用 5 种固定点生存率差别检验法可发现(表2), 两组间第 12 个月和第 18 个月生存率的差异均无统计学意义, 而在第 30 个月时差异均有统计学意义。但是在长期中的第 24 个月时生存率的差异除对数转换法和 logit 转换法无统计学意义外, 其他 3 种方法的结果均有统计学意义, 其中 cloglog 转换法效果最好<sup>[6]</sup>, 已有模拟研究表明该法在第一类和第二类错误上综合效果最佳<sup>[5]</sup>, 本文以 cloglog 的检验效果为准, 可以认为第 24 个月的结果有统计学意义。结果显示在第 18 个月前两组疗效无差异, 而在第 18 个月后万络组的生存率相对较低, 即服用万络患者短期内安全, 但长期服用有风险。

结语: 两条生存曲线比较时, 在满足了两组累积危险率成比例假设的前提下, Log-rank 检验为最佳选择, 否则 Two-stage 检验法更合适, 但有特殊, 如前述 2 个实例, 当难以选择两条生存曲线整体间比较