

· 系列讲座 ·

# 临床流行病学

## 第五讲 时间效应分析(二)

章扬熙

### 四、时间效应数据的比较

1. 时序检验(Logrank test): 是 Mantel N 等人(1966)提出的非参数方法, 它适用于两组或多组时间效应数据的比较。

例 4 有两组麻风病人, 甲组应用多药联合化疗, 乙组应用氨苯砒(DDS)治疗。两组病人的治疗病期随访记录如下:

甲组: 2 3 9 10 10 12<sup>+</sup> 15 15<sup>+</sup> 16  
18<sup>+</sup> 24<sup>+</sup> 30 36<sup>+</sup> 40<sup>+</sup> 45<sup>+</sup>

乙组: 9 12<sup>+</sup> 16 19 19<sup>+</sup> 20<sup>+</sup> 20<sup>+</sup>  
24<sup>+</sup> 24<sup>+</sup> 30<sup>+</sup> 31<sup>+</sup> 34<sup>+</sup> 42<sup>+</sup> 44<sup>+</sup> 53<sup>+</sup>  
59<sup>+</sup> 62<sup>+</sup>

计算步骤详见表 4。表 4 的第(1)、(2)、(3)、(4)、(5)栏的计算过程与例 1 的 Kaplan Meier 法的开始部分相同。第(6)栏甲、乙两组的合并治愈率用下式计算:

$$q_i = \frac{h_{甲} + h_{乙}}{n_{甲} + n_{乙}} \quad (12)$$

表 4 Logrank 检验的计算

治疗病期 $t_i$	期初病例数		$t_i$ 时治愈人数		两组合并治愈率 $q_i = \frac{h_{甲} + h_{乙}}{n_{甲} + n_{乙}}$	期望治愈人数	
	$n_{甲}$	$n_{乙}$	$h_{甲}$	$h_{乙}$		$n_{甲} \times q_i$	$n_{乙} \times q_i$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2	15	17	1	0	0.03125	0.47	0.53
3	14	17	1	0	0.03226	0.45	0.55
9	13	17	1	1	0.06667	0.87	1.13
10	12	16	2	0	0.07143	0.86	1.14
15	9	15	1	0	0.04167	0.37	0.63
16	7	15	1	1	0.09090	0.64	1.36
19	5	14	0	1	0.05263	0.26	0.74
30	4	8	1	0	0.08333	0.33	0.67
合计			8	3		4.25	6.75

比如, 第一行  $q_1 = \frac{1+0}{15+17} = 0.03125$ , 余类推。

$$\chi^2 = \frac{(8 - 4.25)^2}{4.25} + \frac{(3 - 6.75)^2}{6.75} = 5.39$$

第(7)栏为甲组期望治愈人数, 用下式计算:  $n_{甲} \times q_i$ , 在第一行该值为  $15 \times 0.3125 = 0.47$ , 余类推。第(8)栏与第(7)栏类似, 只是求乙组期望治愈人数而已, 用公式  $n_{乙} \times q_i$  计算, 在第一行该值为  $17 \times 0.3125 = 0.53$ , 余类推。

查  $\chi^2$  值表,  $df=1$  时,  $\chi_{0.05}^2 = 3.84$ , 今  $\chi^2 > \chi_{0.05}^2$ ,  $P < 0.05$ , 故拒绝两组无差别的检验假设, 甲组疗效优于乙组。

然后, 用下式求  $\chi^2$  值

2. Gehan 检验: 也是一种非参数统计方法, 用于两组时间效应数据的比较。

$$\chi^2 = \sum \frac{(O - E)^2}{E}, df = \text{组数} - 1 \quad (13)$$

例 5 以例 4 资料, 作 Gehan 检验(表 5)。

具体步骤为:

式中 O 为观察值, E 为期望值, 本例

① 检验假设: 两组病人的治疗病期的分布无区别。

② 由于两组来自同一分布, 所以把两组合一起计算。表 5 的第(1)、(2)栏治疗病期及其所属组别由原始数据整理得来。第(3)栏, 如甲组的第一个病

表 5 Gehan 检验的计算

治疗病期 (1)	组别 (2)	治愈时间更短的病人数 (3)	治疗时间更长的病人数 (4)	$u' = (3) - (4)$ (5)
2	甲	0	31	-31
3	甲	1	30	-29
9	甲、乙	2	28	-26
10	乙、甲	4	26	-22
12 <sup>+</sup>	甲、乙	6	0	6
15	甲	6	23	-17
15 <sup>+</sup>	甲	7	0	7
16	甲、乙	7	20	-13
18 <sup>+</sup>	甲	9	0	9
19	乙	9	18	-9
19 <sup>+</sup>	乙	10	0	10
20 <sup>+</sup>	2 乙	10	0	10
24	甲、2 乙	10	0	10
30	甲	10	11	-1
30	3 甲、8 乙	11	0	11

人,其治疗病期为 2 个月,两组共 32 个病人,没有一个病人的治疗病期比它更短,因此记为 0;第二个病人的治疗期为 3,只有第一个病人的治疗病期比它短,所以记为 1,余类推。但要注意计算治疗病期人数时不计算未治愈的截尾数据,比如,15 个月以前有 8 例,但有 2 例是未愈的截尾数据,所以该行第(3)栏的数值是 6,而不是 8。第(4)栏是治疗时间更长的病人数,比如甲组第一个病人,治疗病期为 2 个月,在两组 32 个病人中,除去他本人外,其余 31 人的治疗病期都比他长,所以记为 31,余类推。但要注意这个数据包括该期以后未愈的截尾数据,但不包括该期以前的截尾数据,比如治疗病期为 15 个月一行,其后有 23 个病人,故第(4)栏的值为 23,不包括以前 12<sup>+</sup>,因为 12<sup>+</sup>的具体治疗病期可能比 15 个月短。另外,遇截尾数据行时,该值无法计算,而记为零。余类推。第(5)栏  $u'$  为同一行第(3)栏数值减去第(4)栏数值所得。 $u'$  值表示每个病人治疗病期在全部病人中所占的位置。

③计算甲组或乙组的  $u'$  值之和,用  $u$  表示。本例,甲组  $u$  值为

$$u_{甲} = (-31) + (-29) + (-26) + (-22) + (-22) + 6 + (-17) + 7 + (-13) + 9 + 10 + (-1) + 11 \times 3 = -96$$

同理,乙组的  $u$  值为  $u_{乙} = +96$ ,甲、乙的  $u$  值绝对值相同,正负号不同。

④用下式计算  $u$  值的方差,式中  $n_1$ 、 $n_2$  分别为两组的例数。

$$Var = \frac{n_1 n_2}{(n_1 + n_2)(n_1 + n_2 - 1)} \Sigma(\Sigma u'^2) \quad (14)$$

本例,甲组  $\Sigma u'^2 = 4534$ ,乙组  $\Sigma u'^2 = 2430$ ,

$$Var = \frac{15 \times 17}{(15 + 17)(15 + 17 - 1)} (4534 + 2430) = 1790.14$$

⑤用下式求  $Z$  值

$$Z = \frac{|u|}{\sqrt{Var}} \quad (15)$$

本例,

$$Z = \frac{96}{\sqrt{1790.14}} = 2.26$$

查  $Z$  值 ( $u$  值) 表,  $P = 0.05$  时,  $Z = 1.96$ , 今  $Z > 1.96$ , 所以  $P < 0.05$ , 拒绝作业假设, 说明两组病人的治疗病期有差别, 甲组疗效优于乙组。

**五、Cox 模型:**1972 年英国统计学家 D.E.Cox 提出了一种能处理多因素时间效应分析数据的回归模型——比例危险度模型, 亦称 Cox 模型。

1.Cox 模型的概述:Cox 模型不是直接用生存时间作为方程式的因变量, 而是用危险度 (hazard) 作为因变量, Cox 模型的结构为

$$h(t, x) = h_0(t) e^{\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p} \quad (16)$$

上式中  $h(t, x)$  为危险度, 如果研究对象的结局为死亡, 它则为时刻  $t$  时的死亡率。 $h_0(t)$  为当所有因素 (自变量)  $X_k$  都取 0 值时的危险度, 它是基准危险度。 $\beta_1, \beta_2, \dots, \beta_p$  是 Cox 模型的回归系数, 它是需从样本数据作出估计的参数,  $k = 1, 2, \dots, p$ 。从上式可见, 各自变量  $X_i$  的回归系数  $\beta_i$  与危险度间呈指数关系, 这是 Cox 模型的一个假定条件。当回归系数  $\beta_k$  等于零时, 说明因素  $X_k$  对危险度不起作用;  $\beta_k$  为正值时  $X_k$  为危险因子, 增大了危险度;  $\beta_k$  为负值时,  $X_k$  为保护因子, 缩小了危险度。当  $X_k$  为 (0, 1) 型变量时, 把  $X_i = 1$  与  $X_k = 0$  的两个危险度相比, 则得到一个与  $h_0(t)$  无关的比值, 即相对危险度。

例 6 麻风病的发病, 与家庭接触瘤型麻风病人  $X_1$  和居住条件拥挤  $X_2$  有关, 对有关资料作 Cox 回归分析后, 得  $X_1$  的回归系数  $\beta_1$  为 0.8755, 得  $X_2$  的回归系数  $\beta_2$  为 0.5108,  $X_1$  与  $X_2$  都是 (0, 1) 变量,  $X_1 = 1$  为有家庭接触瘤型麻风病人情况,  $X_1 = 0$  为无此情况;  $X_2 = 1$  为居住拥挤,  $X_2 = 0$  为居住不拥挤, 试求既接触瘤型麻风又居住拥挤与无接触麻风又居住不拥挤的危险度比值 (即比例危险度)。

将原题给出的已知数据代入公式 (16) 得既不接触瘤型麻风又居住不拥挤的危险度 =  $h_0(t)$

$$e^{(0.8755 \times 1 + 0.5108 \times 1)} = 4ho(t)$$

既不接触瘤型麻风又居住不拥挤的危险度 =  $ho(t)$

$$e^{(0.8755 \times 0 + 0.5108 \times 0)} = ho(t)$$

$$\text{二者的比值} = \frac{4ho(t)}{ho(t)} = 4$$

说明前者危险度是后者的 4 倍。

关于 Cox 模型的参数估计,由于该模型中  $ho(t)$  未作定义,它属半参数模型,故先构成偏似然函数后,应用最大似然法来估计  $\beta_k$ 。对 Cox 模型的选择,常采用后退法,即先把全部因子引入模型,然后再通过假设检验把无显著性的因子剔除。常用的参数检验的方法有梯度检验, Wald 检验(即  $u$  检验)及似然比检验,在大样本中三者的结论是一致的。

对第  $j$  个研究对象的危险(预后)指数由下式求得

$$W_j = \sum_{k=1}^p [\beta_k (X_{jk} - \bar{X}_k) / S_k] \quad (17)$$

上式中  $\bar{X}_k$  与  $S_k$  为均数和标准差,  $(X_{jk} - \bar{X}_k) / S_k$  为  $X_{jk}$  的标准化值。危险指数反映该研究对象的危险程度(预后好坏)。

为了研究多因素对生存率(效应率)的影响,需把实际资料分成若干组,求各组的累计生存率(效应率)(或累计发病率等)。并绘制有关曲线。可按危险指数分组,以各组段组中值计算生存率(效应

率),来考察不同危险指数的生存率(效应率)的差异。也可以按因素的状态分组,以考察不同因素及不同因素组合的生存率(效应率)。生存率(效应率)用下式计算

$$S(t, x) = [So(t)]^{exp(\beta X_i)} \quad (18)$$

上式中  $So(t)$  为基准生存(效应)函数,可用下式求得

$$So(t) = exp[-Ho(t)] \quad (19)$$

上式中  $Ho(t)$  为基准累积危险函数,用下式估计

$$Ho(t) = \sum_{i=ii < t} [di / \sum_{j \in Ri} exp(\beta X_j)] \quad (20)$$

2. Cox 模型的适用条件及在临床流行病学中的应用: Cox 模型适用于处理单因素或多因素影响下的时间——效应数据,这种数据除包括每个研究对象的有关因素外还有生存时间,或结局未发生的截尾数据。应当指出,只有截尾数据是不能进行时间效应分析的,必须有结局发生的生存时间数据。在临床流行病学的定群研究与随访研究中,常用 Cox 模型作为统计分析的手段;尤其是在考察远期治疗效应方面,它在研究设计上比用 logistic 模型要灵活些。比如,对随访迟早不一、随访时间长短不一情况善于处理,对失访情况的处理也较容易。

3. Cox 模型计算要应用 Cox 模型专用程序在电子计算机上运算,程序的使用方法可参考有关的使用说明。  
(收稿:1996-09-15)

## 2724 名小学生乙型肝炎血源疫苗免疫效果观察

刘占文 王玉霞

为了解乙型肝炎血源疫苗(以下简称乙肝疫苗)免疫效果,1995 年 4 月对牡丹江市 2724 名小学生免疫情况进行了调查,现报告如下。

一、对象与方法:对 2724 名小学生进行乙肝疫苗免疫。采用 0、1、6 免疫程序,剂量  $10\mu\text{g}/\text{次}$ ,部位前臂三角肌,有免疫史者 0 月加强一针,剂量  $10\mu\text{g}$ 。全程免疫两周后进行抗-HBs 检测。乙肝疫苗是上海实业科华生物技术有限公司生产,批号:940918。乙型肝炎表面抗体诊断血球是上海生物制品研究所生产,批号:931201-6。

二、结果和分析:2724 人中抗-HBs 阳性 2579 人,占 94.68% (2579/2724),抗-HBs 阴性 145 人,占 5.33% (145/2724)。其中,首次免疫者 2362 人,抗-HBs 阳性者 2219 人,占 93.95% (2219/2362),

抗-HBs 阴性者 143 人,占 6.05% (143/2362);有免疫史者 362 人,抗-HBs 阳性者 360 人,占 99.45% (360/362),抗-HBs 阴性者 2 人,占 0.55% (2/362)。首次免疫抗-HBs 阳性 2219 人中,抗-HBs 滴度在 1:16 以下 273 人,占 12.3% (273/2219),滴度在 1:16 以上 1946 人,占 87.7% (1946/2219);有免疫史者抗-HBs 阳性 360 人中,抗-HBs 滴度 1:16 以下 18 人,占 5% (18/360),滴度 1:16 以上 342 人,占 95% (342/360)。

首次免疫者抗-HBs 阳性率可达 93.95%,滴度 1:16 以上达 87.7%,证明我国乙肝疫苗免疫效果很好。加强免疫者抗-HBs 阳性率达 99.45%,滴度 1:16 以上达 95%,说明及时加强免疫能提高乙肝疫苗免疫效果。

(收稿:1996-09-03 修回:1996-10-25)