·系列讲座·

肿瘤流行病学研究资料的统计分析

项永兵

第六讲　相对生存率的统计学检验

相对生存率定义为研究人群的观察生存率与假定其为一般人群时的期望生存率之比，校正了诸如性别、年龄、年代等因素对研究人群生存情况的影响[1, 2]。在分析比较时，两组或多组研究人群相对生存率的估计可以采用多种方法。但它们各自所代表的总体相对生存率是否相等，这是研究者经常需要回答的问题。本讲将结合实例分析就相对生存率假设检验的若干问题及统计推断作一介绍。

一、假设检验的建立及统计推断：样本相对生存率是总体相对生存率的点估计，由于存在抽样误差，故需进行假设检验。常用的检验方法有下面几种。

1. Hakulinen 氏似然比检验[3]：

①无效假设：样本(m)来自一个相同的总体，具有相同的区间(i)别相对生存率，即 H_0：$r_{ki}=r_{1i}$，$k=2\cdots, m$；$i=0, \cdots, w-1$。

②备择假设：其一是样本生存率来自不同的总体，同时满足 Cox 的比例危险假设[4]，即各样本在随访区间之间均存在一个不随时间而变化的比例危险因子 Ck，即 H_1：$r_{ki}=r_{1k}^{Ck}$，$k=2, \cdots m$；$i=0, \cdots, w-1$；其二是拒绝无效假设 H_0，认为 m 组在所有随访区间中，至少有一个区间与其他区间的相对生存率不等，或 H_2：拒绝 H_0，对 r_{ki}无比例危险假设的限制。

③求似然比统计量：首先假定样本是取自一个参数为(l_{ki}, $1-p_{ki}$)的二项分布总体，则可获得下述对数似然函数：

$$L=\sum_{k=1}^{m}\sum_{i=0}^{w-1}[(l_{ki}-d_{ki})\ln(p_{ki}^{*}\times r_{ki})-d_{ki}\times \ln(1-p_{ki}^{*}\times r_{ki})] \qquad (1)$$

其中 l_{ki}为k 组进入区间[i, $i+1$)的随访观察人数，d_{ki}为死于该区间的个体数，p_{ki}^{*}和r_{ki}分别为区间别期望生存率及相对生存率，p_{ki}为相应的观察生存率。其次在 H_2 备择假设情况下，求 r_{ki}的极大似然估计值，$\hat{r}_{ki}=(l_{ki}-d_{ki})/(l_{ki}\times p_{ki}^{*})$，代入上述(1)式，求似然函数的极大值，作为 L_2。然后在 H_0 无效假设情况下，对 r_{1i} 求一阶偏导数并使之等于零，即 $\partial Ln(L)/\partial r_{1i}=0$($i=1, \cdots, g$)。从而得到下式：

$$\sum_{k=1}^{m}Ck\left[l_{ki}-\frac{d_{ki}}{1-p_{ki}^{*}(r_{1i})^{Ck}}\right]=0,\ i=1, \cdots, g。 \qquad (2)$$

在 H_1 备择假设情况下，一方面要对 r_{1i}求一阶偏导使之等于零；同时还要对 Ck 求一阶偏导使之等于零，即 $\partial Ln(L)/\partial r_{1i}=0$($i=1, \cdots, g$)和 $\partial Ln(L)/\partial Ck=0$($k=1, \cdots, m$)。这样得到(2)式及下式：

$$\sum\ln r_{1i}\left[l_{ki}-\frac{d_{ki}}{1-p_{ki}^{*}(r_{1i})^{Ck}}\right]=0,\ k=2, 3, \cdots, m \qquad (3)$$

通过用 Newton－Raphson 迭代求解(2)式获得 L 的极大值 L_0，同时求解(2)及(3)式可得到 L 的极大值 L_1。

有了不同假设检验情况下的似然函数极大值，就可以求下述似然比卡方统计量。检验 H_0 对 H_1：$\chi_1^2=-2(L_0-L_1)$，$df=m-1$；检验 H_0 对 H_2：$\chi_T^2=-2(L_0-L_2)$，$df=g(m-1)$；检验 H_1 对 H_2：$\chi_2^2=-2(L_1-L_2)$，$df=m-1$。很显然，$\chi_T^2=\chi_1^2+\chi_2^2$。

需要说明的一点是，当区间[i, $i+1$)内存在终检个体 w_{ki}时，可以用 l'_{ki}取代 l_{ki}代入(1)式，其中 $l'_{ki}=l_{ki}-\frac{w_{ki}}{2}$。

2. 比分检验：Brown 氏比分检验[5]主要用于两样本相对生存率差别的比较。通过计算下述卡方统计量进行统计推断，

$$\chi_B^2=S^2/D^2S, \qquad (4)$$

式中 $S=\sum_{i=0}^{w-1}\frac{1-\hat{r}_{1i}}{1-\hat{r}_{1i}\times p_{1i}^{*}}\quad d_{1i}-(1-\hat{r}_{1i}\times p_{1i}^{*})l_{1i}$，

作者单位：上海市肿瘤研究所流行病学研究室 200032

$$D^2S=\sum_{i=0}^{w-1}\frac{l_{1i}\times l_{2i}\times p_{1i}^*\times p_{2i}^*\times \hat{r}_{1i}(1-\hat{r}_{1i})^2}{l_{1i}\times p_{1i}^*(1-\hat{r}_{1i}\times p_{2i}^*)+l_{2i}\times p_{2i}^*(1-\hat{r}_{1i}\times p_{1i}^*)},$$

其中 r_{1i}为 H_0 无效假设情况下的相对生存率估计值。

3. Pearson 型检验[6]:用于两组或多组样本相对生存率间差别的显著性检验。统计量由下式给出:

$$\chi_p^2=\sum_{k=1}^{m}(O_k-E_k)^2\backslash E_k, \qquad (5)$$

式中 $O_k=\sum_{i=0}^{w-1}d_{ki}$, $E_k=\sum_{i=0}^{w-1}(l_{ki}-e_{ki})$,

$$e_{ki}=[(l_{ki}\times p_{ki}^*/\sum_{k=1}^{m}(l_{ki}\times p_{ki}^*)]\times(l_i-d_i),$$

$$l_i=\sum_{k=1}^{m}l_{ki},\quad d_i=\sum_{k=1}^{m}d_{ki}。$$

上述所有假设检验方法在 p_{ki}^*取值为 1 时, 即可用于观察生存率的假设检验。

二、实例应用:本文利用表 1 数据来说明上述假设检验方法的应用。首先在 H_0 及 H_1 假设检验情况下, 把 $Ck=1$ 代入(2)式开始迭代。r_{1i}的初始值取 H_2 假设检验下的 r_{ki}的平均值, 但除外大于等于 0.99的 r_{ki}。如果在迭代过程中出现 $r_{1i}< S_i$($S_i=min_k[(1/p_{ki}^*)^{1/Ck}]$), 则前面一个 r_{1i}及 Si 和的平均值作为 r_{1i}新的迭代值; 如果出现 $r_{1i}\leqslant 0$, 则前面一个 r_{1i}的$\frac{1}{2}$作为 r_{1i}新的迭代值, 继续迭代下去。同时把 r_{1i}代入(3)式迭代。如果 $Ck<0$, 则 Ck 取前面一个 Ck 值的$\frac{1}{2}$, 继续迭代下去。

表 1　某地男性膀胱癌病人年龄组随访资料的分析(1980～1984)

年龄组(岁)	i	l_{ki}	d_{ki}	w_{ki}	p_{ki}	p_{ki}^*	r_{ki}
40～54	1	127	12	1	0.90514	0.98850	0.91567
	2	114	10	0	0.91228	0.98743	0.92389
	3	104	4	2	0.96117	0.98640	0.97442
	4	98	6	19	0.93220	0.98528	0.94613
	5	73	2	17	0.96899	0.98419	0.98456
55～69	1	490	104	0	0.78776	0.96455	0.81671
	2	386	50	0	0.87047	0.96297	0.90394
	3	336	28	6	0.91592	0.96117	0.95292
	4	302	17	57	0.93784	0.95862	0.97832
	5	228	11	69	0.94315	0.95655	0.98599
70～84	1	438	159	0	0.63699	0.90462	0.70415
	2	279	62	0	0.77778	0.90328	0.86106
	3	217	45	4	0.79070	0.89842	0.88009
	4	168	32	30	0.79085	0.89371	0.88490
	5	106	18	33	0.79888	0.88770	0.89994

注: l_{ki}: 期初观察人数, d_{ki}: 区间死亡人数, w_{ki}: 区间终检人数, p_{ki}、p_{ki}^*、r_{ki}: 区间别观察生存率、期望生存率、相对生存率, i: 随访区间

表 2 为男性膀胱癌年龄组之间比较的结果。从结果来看, 按 α=0.001水准拒绝无效假设 H_0, 接受备择假设 H_1 及 H_2, 可以认为各年龄组在每个随访区间的相对生存率不同, 差异有非常显著性; 同时按 α=0.05水准拒绝 H_1 接受 H_2, 即各年龄组在随访区间之间均存在"比例危险因子"。

表 2　不同年龄组间相对生存率差别的似然比检验

检验假设	统计量	自由度(df)	P 值
H_0 对 H_1	$\chi_1^2=38.611$	2	<0.001
H_1 对 H_2	$\chi_2^2=9.071$	8	>0.05
H_0 对 H_2	$\chi_T^2=47.682$	10	<0.001

表 3 为观察生存率的假设检验结果, 同样可以认为 3 组病人的观察生存率间也存在非常显著的差别。表 4 是不同假设检验情况下, 各比较组的区间别相对生存率的估计值。无效假设 H_0 实际上就是假设它们来自 40～84 岁年龄合计的总体, 具有相同的区间别相对生存率及观察生存率。H_2 假设下的极大似然估计值无需迭代, 区间别生存率即是由样本所获得的观察值, 见表 1。在 H_0 及 H_1 假设下则分别通过 Newton—Raphson 迭代获得 r_{ki}的极大似然估计值。

表 3　不同年龄组间观察生存率差别的似然比检验

检验假设	统计量	自由度(df)	P 值
H_0 对 H_1	$\chi_1^2=120.190$	2	<0.001
H_1 对 H_2	$\chi_2^2=10.761$	8	>0.05
H_0 对 H_2	$\chi_T^2=130.951$	10	<0.001

表 4 3 个年龄组在不同假设情况下区间别相对生存率的估计值

区间	年龄合计	H_0			40~54		55~69		70~84	
	40~84	40~54	55~69	70~84	H_1	H_2	H_1	H_2	H_1	H_2
1	0.79146	0.79146	0.79146	0.79146	0.89175	0.91567	0.82653	0.81671	0.70034	0.70415
2	0.89653	0.89653	0.89653	0.89653	0.94528	0.92389	0.91066	0.90394	0.83950	0.86106
3	0.94302	0.94302	0.94302	0.94302	0.96665	0.97442	0.94515	0.95292	0.89991	0.88009
4	0.95382	0.95382	0.95382	0.95382	0.97198	0.94613	0.95384	0.97832	0.91545	0.88491
5	0.97520	0.97520	0.97520	0.97520	0.98223	0.98456	0.97062	0.98599	0.94578	0.89994

注: 迭代次数: H_0: 21, H_1: 106

三、结语: 从理论上讲, 极大似然比检验有其一定的优越性。参数估计比较稳定, 可靠性较好。Brown 氏比分卡方统计量的推导有赖于它的无效假设: 常数比数比[5], 即 $H_1^{'} R_k = [r_{ki}/(1-r_{ki})]/[r_{1i}/(1-r_{1i})]$。当随访区间间距较小时, $r_{ki}/r_{1i} \approx 1$, 则 $R = \frac{(1-r_{1i})}{(1-r_{ki})}$。在 H_1 假设检验情况下, $R_k = 1/Ck$。而且, 生存分析中比例危险假设的应用似乎比常数比数比假设广泛的多。此外, Brown 氏检验的无效假设存在一些缺陷: ① 取决于随访期中各随访区间长度的大小, ② 当 $r_{ki} = 1$ 时, R_k 不存在, ③ 当 $r_{ki} > 1$ 时, R_k 变为负值; 其次, Brown 氏卡方统计量 χ_B^2 与 Hakulinen 氏卡方统计量 χ_1^2 仅在两样本比较时接近。当样本数增加的时候, Brown 氏卡方统计量的计算相当困难[3]。Buckley 氏 Pearson 型卡方统计量虽然计算容易, 但似乎过于保守, 统计效率较低[3]。这种保守性在样本中带有较多的重合(tie)或终检数据时表现的尤为突出[3,7]。因此, Hakulinen[3] 建议在大样本数据分析时尽量避免采用此法。从表 5、6 可以看出, 几种检验方法的结果类似。尤其是 Brown 氏统计量 χ_B^2 与 Hakulinen 氏统计量 χ_1^2 比较接近。而 Buckley 氏 Pearson 型卡方统计量始终低于前两者, 无论是观察生存率还是相对生存率, 结果均是如此。

表 5 不同假设检验方法用于相对生存率差别比较的统计量及显著性水平

(40~59 岁与 60~79 岁两组)

	χ_1^2	χ_2^2	χ_T^2	χ_B^2	χ_P^2
统计量	30.607	0.268	30.875	28.725	22.683
自由度	1	4	5	1	1
P 值	<0.001	>0.1	<0.001	<0.001	<0.001

表 6 不同假设检验方法用于观察生存率差别比较的统计量及显著性水平

(40~59 岁与 60~79 岁两组)

	χ_1^2	χ_2^2	χ_T^2	χ_B^2	χ_P^2
统计量	71.385	0.158	71.543	64.876	53.003
自由度	1	4	5	1	1
P 值	<0.001	>0.1	<0.001	<0.001	<0.001

当上述所有检验方法用于观察生存率的假设检验时, 此时的(4)式与 Mantel 氏 logrank 检验[8] 类似, 而(5)式则与 Peto 和 Pike 的 logrank 检验[9] 很接近。这在前面的第三讲中有过详细介绍。

比较两个或多个样本生存期差别, 就是各样本不同时点生存率的综合分析。在这种时候研究者常常要考虑各样本内部构成的均衡性问题。所以在样本间各随访时点观察生存率的综合比较时, 如果存在混杂因素, 常常采用分层分析技术计算调整了某因素的卡方统计量做统计推断。由于相对生存率校正了性别、年龄等因素的作用, 所以样本间相对生存率差别的显著性检验是否需要调整性别、年龄等混杂因素的作用, 文献中没有加以讨论。问题是如果样本间除了性别、年龄构成不同, 还存在其他混杂因素, 如何调整, 具体采取什么方法, 这些都是有待于探讨的问题。当然可以在多变量回归模型分析中解决上述问题。有关相对生存率的回归模型分析, 笔者将另文介绍。

(参考文献备索)