

# 晚期非小细胞肺癌 Buckley-James 模型生存分析

徐添 赵洪瑜 严煜 王苏容 朱湘竹

**【摘要】** 目的 探讨影响晚期非小细胞肺癌患者生存时间的因素并建立生存时间预测模型。方法 分析 2004—2006 年南通大学附属医院初治的 184 例晚期非小细胞肺癌患者的临床资料,采用 Buckley-James 生存分析(B-J)模型对可能影响预后的因素进行筛选,运用 log-rank 检验对患者的实际生存时间与预测时间进行比较。结果 B-J 模型因素筛选结果显示,KPS 评分、临床分期、治疗方式以及治疗前血红蛋白水平是晚期非小细胞肺癌患者生存时间的主要影响因素,回归方程为  $\ln\text{MONTH} = 0.0108 \text{ KPS 评分} + 0.0238 \text{ 血红蛋白} + 0.4614 \text{ III b 期} + 0.8027 \text{ III a 期} + 0.3869 (\text{放疗} + \text{化疗}) + 0.507 (\text{放疗} + \text{手术}) + 0.6082 (\text{化疗} + \text{手术}) - 2.098$ 。所建立的生存时间预测模型与实际生存时间比较,差异无统计学意义( $P = 0.575 > 0.05$ )。结论 KPS 评分、临床分期、治疗方式以及治疗前血红蛋白水平均可能影响患者的预后,所建立的生存时间预测模型具有实际意义。

**【关键词】** 晚期非小细胞肺癌; 模型; Buckley-James; 生存时间

**Survival analysis on advanced non-small cell lung cancer with a Buckley-James model** XU Tian<sup>1</sup>, ZHAO Hong-yu<sup>2</sup>, YAN Yu<sup>3</sup>, WANG Su-rong<sup>4</sup>, ZHU Xiang-zhu<sup>1</sup>. 1 *Epidemiology and Health Statistics Department of Public Health College of Nantong University, Nantong 226001, China;* 2 *Department of Radiotherapy;* 3 *Department of Cardiothoracic Surgery, Affiliated Hospital of Nantong University;* 4 *Nursing School of Nantong University*  
Corresponding author: ZHU Xiang-zhu, Email: zhxzh@ntu.edu.cn

This work was supported by a grant from the Health Department of Jiangsu Province (No. H200924).

**【Abstract】** **Objective** To analyze the risk factors related to survival time of advanced non-small cell lung cancer (NSCLC) and to establish a prediction model on survival time. **Methods** From 2004—2006, 184 patients with advanced NSCLC were enrolled in the Affiliated Hospital to the Nantong Medical College. Related risk factors were analyzed, using the Buckley-James model. Both actual and predicted survival time were compared by log-rank test. **Results** Through Buckley-James model analysis, data showed that KPS, clinical stage, treatment and pre-treatment hemoglobin were main influencing factors on survival time. Regression equation appeared to be  $\ln\text{MONTH} = 0.0108 \text{ KPS} + 0.0238 \text{ HB} + 0.4614 \text{ III b} + 0.8027 \text{ III a} + 0.3869 (\text{radiotherapy} + \text{chemotherapy}) + 0.507 (\text{radiotherapy} + \text{operation}) + 0.6082 (\text{chemotherapy} + \text{operation}) - 2.098$ . There was no statistical difference between the prediction and the actual models of survival time by log-rank test ( $P = 0.575 > 0.05$ ). **Conclusion** KPS, clinical stage, treatment and pre-treatment hemoglobin might be associated. Both the prognosis of patients with advanced NSCLC and the prediction model seemed to have practical significances.

**【Key words】** Advanced non-small cell lung cancer; Buckley-James model; Survival time

肺癌是世界最常见的恶性肿瘤之一,近年来我国肺癌的发病率和死亡率一直呈明显上升趋势,已在所有恶性肿瘤中居首位。预计到 2025 年,我国每

年死于肺癌的人数将接近 100 万。非小细胞肺癌(NSCLC)约占肺癌的 80%,确诊时 70% 为晚期(III、IV)病变<sup>[1]</sup>。目前,运用 Cox 风险比例回归模型探讨晚期 NSCLC 预后影响因素的研究已较普遍,但研究结果尚不统一且往往忽视了该模型“比例风险”假定的验证。本研究利用 Buckley-James (B-J) 模型对晚期 NSCLC 患者进行生存分析,进而为该病预后的判断、治疗方案的选择以及生存时间的预测提供依据。

DOI:10.3760/cma.j.issn.0254-6450.2010.10.024

基金项目:江苏省卫生厅研究课题(H200924)

作者单位:226001 南通大学公共卫生学院流行病与卫生统计学教研室(徐添、朱湘竹);南通大学附属医院肿瘤放疗科(赵洪瑜),胸外科(严煜);南通大学护理学院(王苏容)

通信作者:朱湘竹, Email: zhxzh@ntu.edu.cn

### 对象与方法

1. 研究对象:来自南通大学附属医院肿瘤放疗科 2004—2006 年初治的晚期 NSCLC 患者。入选标准:经组织学或细胞学确诊并进行全面分期检查(胸部 CT, 上腹部 B 超或 CT, 骨扫描等)的晚期 NSCLC 患者(Ⅲ或Ⅳ期)。共入组 184 例,随访截止时间为 2010 年 2 月,无失访病例。

2. 随访与观察指标:采用生存期为随访指标。生存期定义为从确诊为晚期 NSCLC 到死亡的时间。死亡是指由于肺部恶性肿瘤生长与扩散而引起的一系列临床综合征所致。其他未知原因的死亡视为截尾数据,随访结束时未发生死亡的也定义为截尾数据,死亡的判定以研究对象的病例记录和死亡通知书为依据,不能取得者,由研究对象的家属提供详细死亡情况。

在研究对象的住院病历中查阅可疑影响因素,包括年龄、性别、病理类型(鳞癌、腺癌、腺鳞癌)、KPS 评分、临床分期(按国际 TNM 分期法分为Ⅲa 期、Ⅲb 期、Ⅳ期)、治疗方式(单纯放疗,放疗+化疗、放疗+手术、化疗+手术)以及血红蛋白(HB)、血清乳酸脱氢酶(LDH)、血小板计数(PLT)、癌胚抗原(CEA)、CAI-99 五种指标的治疗前水平。

3. 指标的量化:由于 B-J 模型为线性模型,为此对所有指标进行量化处理。其中,年龄、KPS 评分、HB、LDH、PLT、CEA、CAI-99 为计量指标,可以直接进入模型。性别为二分类指标(用 0, 1 表示,男性=1,女性=0)。病理类型、临床分期、治疗方式为多分类变量,需要进行哑变量处理后方可进入模型。

4. 统计学分析:B-J 模型为一种处理右删失数据的线性模型,可以针对所有可能影响预后的因素进行概括性分析与列线图分析,从而得到各因素的不同水平对生存时间影响的差异。在列线图中,通过每个因素的取值能够得到一个得分(points),再将得分相加得到一个总的得分(total points),根据总得分可预测其生存时间。由于生存时间一般不是正态分布,常为正偏态分布,故对生存时间进行以 e 为底的对数变换,经过 B-J 函数将全部可能的危险因素进行分析,筛选出对生存时间具有显著性影响的因素(P<0.05),建立对数生存时间与因素之间的线性回归模型。将模型所预测的患者生存时间与其实际生存时间做 log-rank 检验,以证明模型的有效性。

所有资料采用奥克兰大学统计学系 Ross Ihaka 和 Robert Gentleman 共同创立的免费统计软件 R 进

行分析<sup>[2]</sup>。

### 结 果

1. 一般情况:184 例随访病例中,男性 125 例,女性 59 例,男女之比 2.12:1;最大 87 岁,最小 36 岁,平均年龄(63.9±11.2)岁。Ⅲ期患者 126 例,Ⅳ期患者 58 例,Ⅲ、Ⅳ期患者比例为 2.17:1。1、3、5 年的生存率分别为 53%、12%、5%。

2. 单因素分析:对 184 例晚期 NSCLC 患者的年龄、性别、病理类型、KPS 评分、临床分期、治疗方式以及 HB、LDH、PLT、CEA、CAI-99 五种指标的治疗前水平进行 log-rank 检验,在 α=0.05 的水平上发现病理类型、KPS 评分、临床分期、治疗方式、治疗前 HB 水平、CAI-99 与患者生存时间有关(表 1)。

表 1 各研究因素与晚期 NSCLC 患者生存时间的关系

因 素	χ <sup>2</sup> 值	P 值
年龄	1.208	0.272
KPS 评分	5.360	0.020
HB	7.483	0.006
LDH	0.452	0.501
PLT	0.172	0.678
CEA	0.193	0.660
CAI-99	5.959	0.015
性别	1.259	0.262
男性=1		
女性=0		
病理类型	7.380	0.025
腺癌=(0, 0)		
腺鳞癌=(1, 0)		
鳞癌=(0, 1)		
临床分期	15.009	0.001
Ⅳ=(0, 0)		
Ⅲb=(1, 0)		
Ⅲa=(0, 1)		
治疗方式	119.013	0.000
放疗=(0, 0, 0)		
化疗+放疗=(1, 0, 0)		
放疗+手术=(0, 1, 0)		
化疗+手术=(0, 0, 1)		

3. 概括性分析与列线图分析:利用 R 软件的 summary 命令和 nomogram 命令进行概括性分析和列线图分析。概括性分析(表 2)和列线图(图 1)均显示 KPS 评分、HB 与生存时间成正比,年龄、LDH、PLT、CEA、CAI-99 与生存时间成反比,性别与生存时间的关系为:女性>男性,病理类型与生存时间的关系为:鳞癌>腺鳞癌>腺癌,临床分期与生存时间的关系为:Ⅲa 期>Ⅲb 期>Ⅳ期,治疗方式与生存时间的关系为:化疗+手术>放疗+手术>放疗+

表2 各研究因素对NSCLC患者生存时间影响的概括性分析

因素	低	高	差值	效应	SE	95%CI
性别	0	1	1	-0.06	0.14	-0.34 ~ 0.21
生存时间比例	0	1	1	0.94	NA	0.71 ~ 1.23
年龄	55	74.25	19.25	-0.28	0.08	-0.48 ~ -0.11
生存时间比例	55	74.25	19.25	0.72	NA	0.57 ~ 0.90
KPS评分	60	80	20	0.98	0.11	0.73 ~ 1.10
生存时间比例	60	80	20	1.99	NA	1.73 ~ 2.17
HB	109	136	27	1.03	0.05	0.75 ~ 1.33
生存时间比例	109	136	27	2.03	NA	1.75 ~ 2.34
LDH	134	223.75	89.75	-0.49	0.05	-0.55 ~ -0.16
生存时间比例	134	223.75	89.75	0.51	NA	0.45 ~ 0.85
PLT	137.75	233.25	95.5	-0.46	0.07	-0.54 ~ -0.15
生存时间比例	137.75	233.25	95.5	0.54	NA	0.46 ~ 0.86
CEA	4	22	18	-0.48	0.07	-0.61 ~ -0.22
生存时间比例	4	22	18	0.52	NA	0.39 ~ 0.78
CAI-99	7.875	34.25	26.375	-0.71	0.01	-0.80 ~ -0.40
生存时间比例	7.875	34.25	26.375	0.29	NA	0.20 ~ 0.60
病理类型 鳞癌:腺癌	3	1	NA	0.56	0.16	0.20 ~ 0.84
生存时间比例	3	1	NA	1.55	NA	1.22 ~ 1.85
病理类型 腺鳞癌:腺癌	3	2	NA	0.34	0.15	0.05 ~ 0.40
生存时间比例	3	2	NA	1.36	NA	1.06 ~ 1.42
分期 IIIa期:IV期	3	1	NA	1.97	0.44	1.20 ~ 3.14
生存时间比例	3	1	NA	5.41	NA	2.37 ~ 11.41
分期 IIIb期:IV期	3	2	NA	1.02	0.38	0.43 ~ 2.00
生存时间比例	3	2	NA	2.07	NA	1.45 ~ 2.99
治疗方式 化疗+手术:放疗	4	1	NA	1.51	0.71	0.40 ~ 2.31
生存时间比例	4	1	NA	5.18	NA	1.63 ~ 12.77
治疗方式 放疗+手术:放疗	4	2	NA	1.23	0.48	0.28 ~ 2.17
生存时间比例	4	2	NA	4.64	NA	1.57 ~ 12.52
治疗方式 化疗+放疗:放疗	4	3	NA	0.96	0.39	0.19 ~ 2.04
生存时间比例	4	3	NA	4.26	NA	1.27 ~ 9.13

注:NA为未做

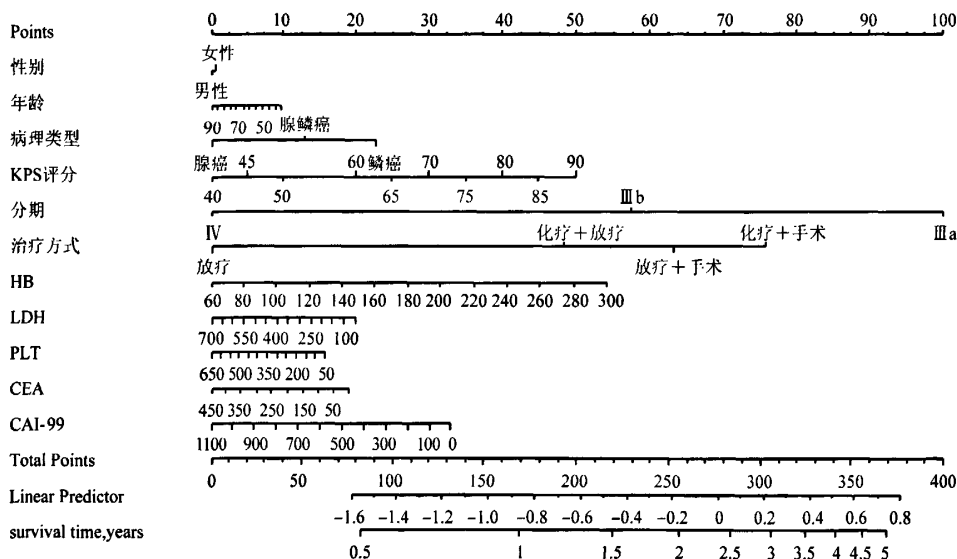


图1 各研究因素对NSCLC患者生存时间影响的列线图分析

化疗>单纯放疗。同时,在列线图中,每个因素向 points 轴作垂线,产生一个交点,能够得到一个得分 (points),再将得分相加得到一个总的得分,根据总

得分可以对其生存时间进行预测。

4. B-J模型因素筛选:经过函数B-J将全部可能的危险因素纳入分析,拟合B-J模型(表3)。结果显

示影响晚期NSCLC患者生存时间的主要因素包括KPS评分、治疗前HB水平、临床分期以及治疗方式,其结果符合临床实际。

表3 拟合B-J回归模型的结果

项目	$\beta$	$s_e$	Z值	P值
常数项	-2.0980	0.6844	-4.0050	0.000 063 2
性别	-0.1872	0.1622	-1.1543	0.973 3
年龄	-0.0014	0.0046	-0.3117	0.755 2
KPS评分	0.0108	0.0054	1.4861	0.013 73
HB	0.0238	0.0015	1.2186	0.022 3
LDH	-0.0002	0.0005	-0.4634	0.643 1
PLT	-0.0002	0.0006	-0.3420	0.732 3
CEA	-0.0003	0.0017	-0.2048	0.837 8
CAI-99	-0.0002	0.0017	-0.1401	0.888 6
腺鳞癌	0.1043	0.1668	0.3267	0.744 0
鳞癌	0.1817	0.1640	1.1558	0.247 8
Ⅲb期	0.4614	0.1639	2.0101	0.044 4
Ⅲa期	0.8027	0.1770	3.8672	0.000 110
化疗+放疗	0.3869	0.2334	1.6579	0.037 34
放疗+手术	0.5070	0.1802	2.8128	0.004 912
化疗+手术	0.6082	0.1936	3.1405	0.001 687

5. B-J模型的建立与检验:根据B-J模型因素的筛选结果建立回归模型: $\ln\text{MONTH}=0.0108 \text{ KPS评分} + 0.0238 \text{ HB} + 0.4614 \text{ Ⅲb期} + 0.8027 \text{ Ⅲa期} + 0.3869(\text{放疗} + \text{化疗}) + 0.507(\text{放疗} + \text{手术}) + 0.6082(\text{化疗} + \text{手术}) - 2.098$ 。根据该模型可以直接对晚期NSCLC患者的生存时间进行预测。同时将B-J模型对本次研究中的184例患者的预测生存时间与其实生存时间做log-rank检验,  $\chi^2=0.3, P=0.575 > 0.05$ 。表明预测生存时间与实际生存时间之间的差异无统计学意义。并做Kaplan-Meier生存曲线(图2),可见两条生存曲线比较靠近,证明该模型有效。

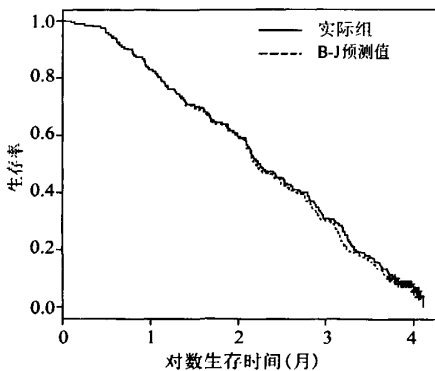


图2 B-J预测生存时间与实际值的比较

6. 多元比例风险模型Cox回归分析:利用Cox多元比例风险模型对可能影响预后的因素进行分析,逐步回归法筛选变量( $\alpha=0.10$ ),最终进入模型

的因素包括KPS评分、临床分期、治疗方式以及治疗前HB水平(表4)。这一结果与B-J模型因素筛选完全一致。

表4 Cox比例风险模型多因素分析

因素	$\beta$	$s_e$	Wald $\chi^2$ 值	P值	RR值(95%CI)
KPS评分	-0.253	0.022		0.002	0.77(0.711 ~ 0.787)
临床分期					
Ⅳ期					
Ⅲb期	-0.499	0.188	7.015	0.008	0.607(0.420 ~ 0.878)
Ⅲa期	-1.671	0.206	65.591	0.000	0.188(0.126 ~ 0.282)
治疗方式					
放疗					
放疗+化疗	-1.001	0.203	24.405	0.000	0.368(0.247 ~ 0.547)
放疗+手术	-1.933	0.261	54.722	0.000	0.145(0.087 ~ 0.241)
化疗+手术	-2.136	0.247	74.726	0.000	0.118(0.073 ~ 0.192)
HB	-0.224	0.018	154.251	0.003	0.799(0.771 ~ 0.828)

### 讨论

本研究利用B-J模型对NSCLC预后影响因素进行分析,结果显示KPS评分、临床分期、治疗方式以及治疗前HB水平是晚期NSCLC患者生存时间的独立影响因素,与国内外研究结果基本一致<sup>[3,4]</sup>。KPS评分反映了患者生活的自理状况,它被纳入模型说明了KPS评分较高的患者对治疗依从性较好,提高了治疗的成效。临床分期与治疗方式两项因素与患者的预后关联性较强,对于晚期NSCLC患者而言,Ⅲa期和部分Ⅲb期患者未发生远隔转移,因此主要采取化疗或放疗缩小病灶后再行手术的方法,有效提高了其生存率;对于已经出现了远隔转移的Ⅲb期与Ⅳ期患者,则根据其身体状况,选择放化疗相结合或姑息放疗的方法。HB浓度直接影响人体血氧含量及肿瘤的供氧,同时氧也是最强的放射增敏剂<sup>[5]</sup>,因此HB含量的高低与患者接受放疗的效果和其生存率均存在密切联系。

根据筛选出的因素所建立的线性回归模型可直接预测晚期NSCLC患者的生存时间。通过log-rank时序检验证明,该模型对本次研究中所有患者生存时间的预测与其实际生存时间的差异无统计学意义,具有良好的预测功能。从Kaplan-Meier生存曲线见两条生存曲线比较靠近,只是到了晚期有些分歧,其原理可能是因为B-J模型的参数估计采用的也是最小二乘法,虽然通过引入伪随机变量,但它只是保证变换后的对数生存时间与预测值之间残差平方和最小,所以导致该结果出现<sup>[6]</sup>。

B-J模型是右删失数据的线性回归模型,其估计方法是普通最小二乘法的一种扩展<sup>[7]</sup>。但需满足以

下条件:①自变量与因变量有线性关系;②方差齐性;③每个自变量至少需要 10~15 个以上的阳性结局事件;④删失比例不超过 60%。其中线性与方差齐性主要是通过残差图进行检验<sup>[8]</sup>。本研究共纳入 184 例患者,随访截止时,尚有 12 例存活,无失访病例,删失比例与样本量均达到了 B-J 模型的要求。根据对数生存时间和各因素作出残差图(图 3)可见,散点围绕以“0”为中心的横轴上下波动,且波动范围不超过“±3”,分布比较均匀,故资料满足了 B-J 模型方差齐性和线性的基本假定。

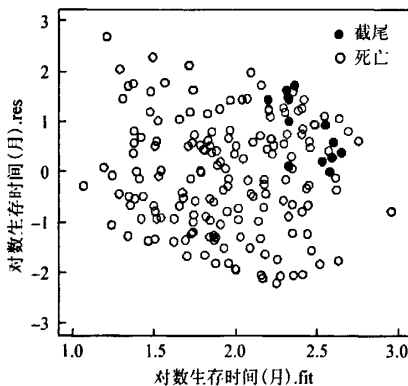


图3 生存时间进行以 $e$ 为底的对数变换后的残差图

同目前已在生存分析领域广泛使用的 Cox 模型相比, B-J 模型具有如下优点:①没有比例风险的假设前提;②该模型属于线性模型,结构简单直观,参数的含义更易于理解;③ B-J 模型可以直接对生存时间进行预测,而 Cox 模型拟合的结果不能用于直接的预测,如果需要预测就必须估计基准风险函数,而基准风险函数的估计是非常困难的<sup>[9]</sup>。由于目前国内流行病学研究中应用较多的统计软件 SAS 和 SPSS 并未集成 B-J 模型的算法,其使用和推广受到了一定限制。本研究所使用的 R 软件作为一款免费统计软件,功能强大,它采用面向对象的统计编程语言,和其他编程语言/数据库之间有很好的接口,在国外流行病学研究领域得到了广泛应用<sup>[10]</sup>。

总之,在生存分析中,选择 B-J 模型还是 Cox 模型应当取决于实际资料的具体情况,不能一概而论<sup>[11]</sup>。从本研究结果来看, KPS 评分、临床分期、治

疗方式以及治疗前 HB 水平是影响晚期 NSCLC 患者生存时间的主要因素,因此,在条件允许的情况下应尽可能选择手术切除为主的治疗方式,同时设法增加 HB 浓度,提高放射治疗的疗效,从而有效提高患者的生存时间。

#### 参 考 文 献

- [1] Walker S. Updates in non-small cell lung cancer. Clin J Oncol Nurs, 2008, 12(4):587-596.
- [2] Crawley MJ. The R Book. London: UK, 2007.
- [3] Song LH, Song XR, Liu MQ, et al. Prognostic factors in patients with stage III and IV non-small cell lung cancer. Chin J Oncol, 2004, 26(6):345-348. (in Chinese)  
宋丽华, 宋现让, 刘美芹, 等. III 期 IV 期非小细胞肺癌预后影响因素分析. 中华肿瘤杂志, 2004, 26(6):345-348.
- [4] Goya T, Asamura H, Yoshimura H, et al. Prognosis of 6644 resected non-small cell lung cancers in Japan: a Japanese lung cancer registry study. Lung Cancer, 2005, 50(2):227-234.
- [5] Wang WD, Sun SP, Wang XD. Influence of peripheral blood hemoglobin concentration in radiotherapy effect of patients with non-small; cell lung carcinoma (NSCLC). Cancer Res Clin, 2004, 16(4):239-241. (in Chinese)  
王卫东, 孙苏平, 王向东. 血红蛋白浓度对非小细胞肺癌放疗疗效的影响. 肿瘤研究与临床, 2004, 16(4):239-241.
- [6] Stare J, Heinzl H, Harrell F. On the use of Buckley and James least squares regression for survival data. New Approaches in Applied Statistics, 2000, 16: 125-134.
- [7] Jin Z, Lin DY, Ying Z. On least-squares regression with censored data. Biometrika, 2006, 93: 147-161.
- [8] Buckley J, James I. Linear regression with censored data. Biometrika, 1979, 66:429-436.
- [9] Stare J, Harrell F, Heinzl H. BJ: an S-Plus program to fit linear regression models to censored data using the Buckley-James method. Computer Methods and Programs in Biomedicine, 2001, 64(1):45-52.
- [10] Chongsuvivatwong V. Analysis of epidemiological data using R and Epicalc(应用 R 软件和 Epicalc 模块程序分析流行病学数据). 蔡乐. 北京: 人民卫生出版社, 2008.
- [11] Xiang YB, Jin F, Gao YT. Application of multiple regression analysis in prognostic factor study on cancer. Cancer, 1992, 12(6):241-245. (in Chinese)  
项永兵, 金凡, 高玉堂. 多元回归模型分析在癌症预后因素研究中的应用. 癌症, 1992, 12(6):241-245.

(收稿日期:2010-03-22)

(本文编辑:张林东)