

发表影像学 SCI 论文常用的统计学方法

李亮, 王良, Jinxing Yu, 冯朝燕, 邓明, 冯定义, 彭莉, 刘继红, 胡道予

【摘要】 随着我国科技工作者对影像学的研究日益深入,在欧美高影响因子期刊发表文章成为国内研究者的共同追求。精确的统计结果是优秀期刊文章的基本需要。但是我国不少影像学论文的统计存在问题。我们总结欧美影像学期刊高质量文章常用的影像统计方法,同时结合多年发表欧美影像期刊的经验简要解析常用影像学统计方法,并引用举例探讨其应用。

【关键词】 统计学;医学;ROC 曲线;McNemar 检验;Levene 检验; t 检验;回归分析

【中图分类号】R195.1; R181.2; R311 **【文献标识码】**A **【文章编号】**1000-0313(2014)06-0718-04

DOI:10.13609/j.cnki.1000-0313.2014.06.036

随着我国对科技创新的投入加大,在美欧影像学杂志上投稿的学术论文日益增多,但是不少中国论文因为统计方法不当而被拒稿。统计学方法作为高质量论文写作的重要组成部分,是否选择及使用恰当的统计学方法将直接导致该文是否发表。为了提高放射科医生在美欧影像学杂志上发表高水平论文,我们对国外影像学杂志中使用频率较高的医学统计学方法进行归纳整理,并列说明,以期能够引起我国影像学界同仁的重视。

ROC 曲线

ROC(receiver operating characteristic)曲线在欧美影像学期刊中是应用最为普及的统计学方法,它是根据一系列不同的二分类方式(分界值或决定阈),以真阳性率(灵敏度)为纵坐标,假阳性率(1-特异度)为横坐标绘制的曲线。1960 年 Lusted 首先将 ROC 分析应用于医学诊断,随后经大量学者的研究和实践,ROC 曲线现已成为广泛应用于临床诊疗和人群筛检研究的常用统计学方法^[1]。它结合灵敏度(sensitivity)和特异度(specificity)对诊断试验进行综合评价,根据曲线的形状和曲线下面积(area under the ROC curve, AUC)的大小对诊断试验作定量分析。理论上 AUC 的取值在 0 和 1 之间。当 $AUC < 0.5$ 时,不符合真实情况在实际中极少出现。在 $AUC > 0.5$ 的情况下, AUC 越接近于 1,说明诊断效果越好。AUC 在 $0.5 \sim 0.7$ 时有较低准确性, AUC 在 $0.7 \sim 0.9$ 时有一定准确性, AUC 在 0.9 以上时有较高准确性。AUC = 0.5 时,说明诊断方法完全不起作用,无诊断价值^[2]。

ROC 分析是影像科研文献中应用最广泛的统计学方法,它常用于选择最佳界值点(cut-off)和比较两种或两种以上影像诊断方法的差异性,因界值点不同而产生动态变化的真阳性率和假阳性率曲线^[3]。Wang 等^[4]在影像科研中使用了 ROC 分析方法,两名放射科医生分别利用磁共振胶片和 PACS 系统两种方法判断前列腺癌包膜侵犯和精囊浸润的情况。评估分为 1~5 五个等级,1 肯定无,2 可能无,3 不确定,4 可能有,5 肯定有。

作者单位:430030 武汉,华中科技大学同济医学院附属同济医院放射科(李亮、王良、冯朝燕、邓明、冯定义、彭莉、胡道予);泌尿外科(刘继红);Department of Radiology, Virginia Commonwealth University, Medical College of Virginia, Richmond, USA(Jinxing Yu, MD)

作者简介:李亮(1983—),男,河北邯郸人,博士研究生,主要从事肿瘤学 MR 研究工作。

通信作者:王良, E-mail: wangliang2001@gmail.com

然后分别以 1、1+2、1+2+3、1+2+3+4 作为阴性诊断指标, 2+3+4+5、3+4+5、4+5、5 作为阳性诊断。计算出每个临界值所对应的特异度和敏感度,并绘制 ROC 曲线。结果发现利用 PACS 系统工具的曲线下面积大于磁共振胶片,说明利用 PACS 系统对前列腺癌包膜侵犯和精囊浸润的诊断价值更大。Chen 等^[5]在研究双反转恢复序列(double inversion)、T₂ FLAIR 序列、增强 FSPGR 及增强 T₂ FLAIR 序列 4 种磁共振成像序列探测脑转移病灶的特异性和敏感性中用到了 ROC 分析。结果发现增强 T₂ FLAIR 是评估脑转移最敏感的序列,它的曲线下面积最大(分别为 0.763、0.709、0.865、0.993)。ROC 曲线也可以用于临界值的估计,如有学者在研究肝周淋巴结转移的实验中,得到门腔淋巴结最大截面积 $> 0.65 \text{ cm}^2$ 认为有淋巴结转移,CT 扫描只要发现胰腺十二指肠淋巴结和肝动脉淋巴结就可以认为有淋巴结转移^[6]。临床影像诊断工作中,检查结果通常并不以二分类的方法出现,所以应首先进行数据转换后方可采用该方法。Wang 等^[7]研究前列腺癌患者盆腔淋巴结转移的单因素和多因素分析结果显示 MRI 可独立地评价前列腺癌淋巴结转移的阳性和阴性预测值分别为 50% 和 96.36%,而 MRI 图像结合 Partin nomograms 方法,其诊断的准确率要高于前者, AUC 的值为 0.892,要高于前者的 0.633。

McNemar 检验

该检验由美国统计学家 Quinn McNemar 在 1947 年提出^[8],用于非参数方法配对计数资料的比较,主要分析配对资料中对照组和实验组的频数或比率是否有差异。也可以分析同一批观察对象影像学检查前后的结果有无差异。配对资料中两种不同影像学处理的实验结果必须是二分类结果如“是或否”、“阳性或阴性”、“有反应或无反应”等。应用条件是配对的样本为随机样本,且对子之间相互独立,资料为二分类结果的计数资料。

McNemar 检验在医学影像学中的应用较为广泛。如 Omoumi 等^[9]采用 CT 和 MR 关节成像两种方法评价肩袖肌腱撕裂患者,McNemar 方法被使用来对比这两种技术在显示肩袖肌病灶的敏感性和特异性。结果表明这两种方法对肩袖肌腱撕裂患者诊断的敏感性和特异性之间无统计学差异,诊断价值相当。Doneda 等^[10]研究产前 MR 在评价巨细胞病毒感染胎儿早期脑损伤的价值。他采用 McNemar 检验的方法对比 MR 和

超声技术在病灶检出率中的差异。结果发现 MR 检测大脑异常和预测感染症状的灵敏度均高于超声成像,分别为 92% vs 38%、83% vs 33%。此外,与超声相比,MR 可以发现更多的脑异常征象,如胎儿脑小畸形,皮质异常、颞极病变等。

Kappa 一致性检验

1960 年 Cohen 等提出用 Kappa 值作为评价一致程度的指标^[11]。Kappa 系数适用于无序分类变量资料,包括两项分类和多项分类资料。在影像学实验中,常用于判断 2 名或多名医师测量同一研究对象或者 1 名医师多次测量同一实验对象的一致性^[12]。若对两种非金标准的诊断方法进行诊断结果一致性分析,可通过计算 kappa 统计量作一致性检验。如果研究者希望考察新的诊断方法与金标准的一致性,则除了作一致性检验以外,还应计算灵敏度、特异度、阳性预测值和阴性预测值等评价指标。Landis 和 Koch 将 Kappa 系数的大小分为 6 个区段^[13],分别代表一致性的强弱程度,即 Kappa 值 <0 时,说明一致性强度极差,0.0~0.2 时为微弱,0.21~0.40 时为弱,0.41~0.60 时为中度,0.61~0.80 时为高度,0.81~1.00 时为极强。

Kappa 一致性检验在影像学专业期刊中应用也比较广泛,如甲乙两名放射科医师根据 MRI 图像判断直肠癌是否浸润直肠壁肌层,分析结果发现($\kappa=0.61$),说明甲乙两医师一致性较高^[14]。两名放射科医师分别用 MRI 图像和 PACS 系统判断前列腺癌分期,kappa 值分别是 0.56 和 0.76,说明两位医师各自具有良好的一致性^[4]。两位观察者分别对 201 例肾脏肿瘤患者是否存在瘤周血管进行判断,结果阅片者 1 诊断 94 例(47%),阅片者 2 诊断 69 例(34%),kappa 值是 0.60,两者具有中等一致性^[15]。

Levene 检验

Levene 检验也称为 Levene 方差齐性检验。由 Levene 首先在 1960 年提出,Brown 和 Forsythe 于 1974 年对 Levene 检验进行了扩展,使其对原始数据的数据转换不但可以使用数据与算术平均数的绝对差,也可以使用数据与中位数和调整均数的绝对差,这就使得 Levene 检验的用途更加广泛^[16]。Levene 检验主要用于检验两个或两个以上样本间的方差齐性,同时要求样本为随机样本且相互独立。Levene 检验既可以用于正态分布的资料,也可以用于非正态分布的资料。Levene 检验法在用于对多总体方差进行齐性检验时,所分析的资料可不具有正态性。

如有学者研究 CT 层厚对肺转移瘤形态测量的影响,10 例患者共计 42 处肺转移病灶,分别以 3.75、5.0 和 7.5mm 层厚进行 CT 重建。然后在不同层厚模式下分别计算每个病灶的最大左右径、前后径、最大横截面积以及体积。结果发现对于直径和面积测量而言,层厚为 5.0 和 7.5mm 并无显著差异,而 3.75mm 时与前两者的测量差异较大^[17]。

χ^2 检验与 Fisher 精确概率法

卡方(χ^2)检验是一种常用的计数资料进行假设检验的统计学方法,主要用于研究两组或多组样本率或构成比之间的差别,两变量间有无关联性以及频数分布的拟合等^[18]。但如果当总例数 $N < 40$,或四格表中最小格子 T 值 < 1 时,就需要采用

Fisher 精确概率法^[19]。此方法不属于卡方检验范畴,但可作为四格表卡方检验应用上的补充。实际上当有统计软件条件下,大样本四格表的资料也可以用 Fisher 精确概率法^[20]。

Mullerad 等^[21]对比泌尿生殖放射科医生和普通放射科医生在评价前列腺癌 MR 包膜外侵犯的差异。344 例前列腺癌患者被分成 2 组,分别由不同专业放射科医生评判。这 2 组间患者的构成比以及病理分期经 χ^2 检验无显著统计学差异(P 值均 > 0.05)。结果表明泌尿生殖放射科医生评价 MR 前列腺癌包膜外侵犯准确率更高。Zhang 和他的同事研究瘤周血管在肾肿瘤评价中的意义。由于部分病理类型肿瘤例数较少,故采用了 Fisher 精确概率法检验。结果发现在各病理亚型肿瘤中,除了肾透明细胞癌外,其它类型肿瘤中发现瘤周血管的比例并无统计学差异^[15]。

t 检验和 Wilcoxon 秩和检验

t 检验是应用最为成熟的参数统计学方法,它分为 3 种,即单组设计定量资料 t 检验、成组设计定量资料 t 检验和配对设计定量资料 t 检验。在影像学杂志中配对设计定量资料 t 检验最为常用。它主要用于评价:①两个同质受检者分别接受两种不同(如 CT 和 MR、MR 和超声等)的处理;②同一受检者接受两种不同的处理;③同一受检者处理前后(如介入治疗前后对比等)。在选用 t 检验时应满足正态性和方差齐性两个前提条件,若资料不能满足两个前提条件解决的办法一是对资料进行相应的变量变换,能满足检验前提条件后,再使用检验;另一是选用适合分析偏态分布的非参数检验(如 Wilcoxon 秩和检验)进行分析^[22-23]。

Wang 等^[24]研究 MRI 上前列腺癌区域 T_2WI 信号与前列腺癌侵袭性间的关系。在文中他使用了 t 检验的方法来评价癌区与非癌区的 T_2 信号强度与 Gleason 评分间的相关性。结果显示 Gleason 评分与前列腺外周带癌区的 T_2 信号强度相关,Gleason 评分越高 T_2 信号强度越低,非癌区域 T_2 信号强度与 Gleason 评分无关。因此 T_2 信号强度可以无创性地预测前列腺癌的侵袭性。Zhang 等^[25]在研究 MR 弥散成像(DWI)评价肾脏病变时用了该检验,他们测量恶性肾肿瘤实性、囊变坏死区域以及良性肾囊肿的表现弥散系数(ADC)值,结果发现肾脏肿瘤的 ADC 值要显著低于肾囊肿(分别为 189.3×10^{-5} 和 $322.8 \times 10^{-5} \text{ mm}^2/\text{s}$, $P < 0.001$);增强扫描肾肿瘤中实性强化部分的 ADC 值亦低于未强化的囊变坏死区(分别为 162.3×10^{-5} 和 $247.7 \times 10^{-5} \text{ mm}^2/\text{s}$, $P < 0.05$)。因此,他认为 DWI 成像将有助于鉴别肾良恶性病灶。此外,国外有学者在研究肾透明细胞癌、乳头状癌、嫌色细胞癌等不同肾脏肿瘤的最大瘤体直径与瘤周血管存在与否的关系时也用了 Wilcoxon 秩和检验,结果发现在不同病理类型肿瘤中,瘤周血管的存在均与肿瘤的体积呈显著相关性,它更多地出现在瘤体较大的恶性肿瘤中,而在相同体积的瘤体中肾透明细胞癌则更频繁可以见到瘤周血管。这种发现将为肾脏恶性肿瘤的术前评价提供更多的影像学诊断信息和预测治疗效果^[15]。

Pearson 相关系数与 Spearman 相关系数

Pearson 相关系数是最常用的评价两个连续变量的相关指

标,又称为积差相关系数,取值-1到1,绝对值越大,说明相关性越强。它要求两个变量均为定量变量,并且属于双变量正态分布资料。该系数的计算和检验为参数方法。Spearman 相关系数又称为秩相关系数,它是利用两变量的秩次大小作线性相关分析,对原始变量的分布不做要求,属于非参数统计方法。因此它的适用范围比 Pearson 相关系数要广的多。即使原始数据是等级资料也可以计算 Spearman 相关系数。对于服从 Pearson 相关系数的数据也可以计算 Spearman 相关系数,但统计效能要比 Pearson 相关系数低一些(不容易检测出两者事实上存在的相关关系)^[26]。

学者研究肾肿瘤的体积、生长情况以及与年龄的关系时采用了 Pearson 相关系数方法。结果发现患者年龄和最初的肿瘤体积并没有明显相关性($P=0.13$),然而年龄与肿瘤倍增时间的倒数(reciprocal of doubling time, RDT)却呈负相关($P=0.03$),每增长 1 岁, RDT 将会下降 3%。此外, RDT 和最初肿瘤体积无显著相关($r=-0.11, P=0.42$)^[27]。Zhou 等^[28]研究轻度脑外伤 1 年后患者脑体积变化以及这些改变与临床症状的相关性。文中 Pearson 相关系数被计算用于评价纵向脑体积变化和神经认知积分之间的关系,而 Spearman 相关系数则用于判断局部脑体积与临床症状(如焦虑、抑郁、脑震荡后遗症等)的相关性。结果表明双侧前扣带回脑白质体积的损失与认知功能测量呈正相关(Pearman 相关系数为 0.65, $P=0.005$)。而在 1 年的随访过程中,左侧扣带回峡部体积与临床焦虑积分则呈负相关(Spearman 相关系数为-0.68, $P=0.007$),与脑震荡后遗症也呈负相关(Spearman 相关系数为-0.65, $P=0.01$)。

Logistic 回归模型与 Cox 回归模型

回归分析是近些年国外高质量影像学杂志中较为常见的统计学方法,一项调查显示在血管介入放射学杂志(Journal of Vascular and Interventional Radiology, JVIR)和放射学(Radiology)杂志中,该方法所占比例分别为 13%和 9%^[29]。而多元回归分析是当今统计学发展的趋势,其中 Logistic 回归模型和 Cox 回归模型应用最广泛。Logistic 回归模型是根据单个或多个连续型或离散型自变量来分析和预测离散型因变量的多元分析方法^[30-31],主要用于危险因素的探索以及预测或判别某病发生的概率。Logistic 回归分析中因变量为二分类或多分类变量,自变量既可以为分类变量,也可以为连续变量。实际中由于二分类更加容易解释,所以二分类的 logistic 回归模型最为常用。Cox 比例风险回归模型(Cox's proportional hazards regression model)简称 Cox 回归模型。该模型由英国统计学家 Cox 于 1972 年提出,主要用于肿瘤和其它慢性病的预后分析,也可用于队列研究的病因探索^[32]。两者都属于多元统计分析范畴。区别在于当你只讨论影响结局是否发生的因素时,选用 Logistic 回归。若你不仅关心结局,而且还关心结局发生所经历的时间时,则多因素分析采用 Cox 回归分析。

Wang 等^[33]应用直肠内线圈高场磁共振预测前列腺癌精囊腺侵犯时,分别使用了单因素和多因素 Logistic 回归模型。结果显示在单因素回归分析时, MRI 图像积分($P<0.001$)、血清 PSA 值($P=0.014$)、Gleason 评分($P=0.01$)、临床分期($P<0.001$)、前列腺癌灶的最大比率($P<0.001$)以及神经丛侵犯

($P=0.005$)均与前列腺癌精囊腺侵犯有关,而与阳性穿刺的百分率($P=0.30$)无关。然而在进一步的多因素回归分析时,发现 MRI 图像积分($P<0.001$)、血清 PSA 值($P=0.012$)、Gleason 评分($P<0.001$)及前列腺癌灶的最大比率($P=0.02$)均与前列腺癌精囊腺侵犯显著有关。

Michael 等^[34]在研究 MRI 预测前列腺癌根治性切除术后复发中的作用时,则用到了 Cox 比例风险回归模型。结果前列腺癌术后 5 年复发的模型中,血清 PSA 值($P<0.001$)、临床分期($P=0.006$)、初次 Gleason 评分($P<0.001$)、二次 Gleason 评分($P=0.008$)以及 MRI 图像积分($P<0.001$)均与前列腺癌术后复发有关,而与活检 Gleason 评分总和($P=0.179$)无关。MRI 图像积分与癌灶复发的危险比(Hazard ratio, HR)为 1.76。而前列腺癌术后 10 年复发的模型则显示血清 PSA 值($P<0.001$)、临床分期($P=0.005$)、初次 Gleason 评分($P=0.008$)、阳性穿刺的次数($P=0.009$)以及 MRI 图像积分($P<0.001$)均与前列腺癌术后复发有关,而与二次 Gleason 评分($P=0.067$)和阴性穿刺的次数($P=0.565$)无关。MRI 图像积分与癌灶复发的 HR 为 1.81。这种发现表明 MRI 图像积分系统将会成为前列腺癌根治性切除术后复发的一种有效预测工具。

事实上,本文中仅提到了常用影像统计学方法的一点皮毛,而在日常工作中常面对的是多因素的复杂统计资料,不存在利用某一种统计方法回答专业上期待解决的全部问题。因此,这就需结合专业和统计学知识,选择不同的变量子集,进行各种相应的统计分析,使专业知识和统计知识密切结合,对结果做出合理的分析与解释。

综上所述,以上是我们在欧美放射学杂志上发表论文中关于统计学方法的一点体会。只要我们重视医学论文统计学方法的学习,打下坚实的统计学基础,就能减少统计学上的失误,从而让国外高水平的影像学期刊收录更多来自中国的文章,增加我国放射学从业人员在国际上的影响力和知名度。

参考文献:

- [1] Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests[J]. Prev Vet Med, 2000, 45(1-2): 23-41.
- [2] Park SH, Goo JM, Jo CH. Receiver operating characteristic (ROC) curve; practical review for radiologists[J]. Korean J Radiol, 2004, 5(1): 11-18.
- [3] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve[J]. Radiology, 1982, 143(1): 29-36.
- [4] Wang L, Zhang J, Schwartz LH, et al. Incremental value of multiplanar cross-referencing for prostate cancer staging with endorectal MRI[J]. AJR, 2007, 188(1): 99-104.
- [5] Chen W, Wang L, Zhu W, et al. Multicontrast single-slab 3D MRI to detect cerebral metastasis[J]. AJR, 2012, 198(1): 27-32.
- [6] Grobmyer SR, Wang L, Gonen M, et al. Perihaptic lymph node assessment in patients undergoing partial hepatectomy for malignancy[J]. Annals of Surgery, 2006, 244(2): 260-264.
- [7] Wang L, Hricak H, Kattan MW, et al. Combined endorectal and phased-array MRI in the prediction of pelvic lymph node metastasis in prostate cancer[J]. AJR, 2006, 186(3): 743-748.

- [8] Mc NQ. Note on the sampling error of the difference between correlated proportions or percentages[J]. Psychometrika, 1947, 12(2):153-157.
- [9] Omoumi P, Bafort AC, Dubuc JE, et al. Evaluation of rotator cuff tendon tears: comparison of multidetector CT arthrography and 1.5T MR arthrography[J]. Radiology, 2012, 264(3):812-822.
- [10] Doneda C, Parazzini C, Righini A, et al. Early cerebral lesions in cytomegalovirus infection: prenatal MR imaging[J]. Radiology, 2010, 255(2):613-621.
- [11] McHugh ML. Interrater reliability: the kappa statistic[J]. Biochem Med (Zagreb), 2012, 22(3):276-282.
- [12] Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation and sample size requirements[J]. Phys Ther, 2005, 85(3):257-268.
- [13] Landis JR, Koch GG. The measurement of observer agreement for categorical data[J]. Biometrics, 1977, 33(1):159-174.
- [14] Vliegen RF, Beets GL, von Meyenfeldt MF, et al. Rectal cancer: MR imaging in local staging——is gadolinium-based contrast material helpful? [J]. Radiology, 2005, 234(1):179-188.
- [15] Zhang J, Lefkowitz RA, Wang L, et al. Significance of peritumoral vascularity on CT in evaluation of renal cortical tumor[J]. J Comput Assist Tomogr, 2007, 31(5):717-723.
- [16] Hines WG, Hines RJ. Increased power with modified forms of the Levene (Med) test for heterogeneity of variance[J]. Biometrics, 2000, 56(2):451-454.
- [17] Zhao B, Schwartz LH, Moskowitz CS, et al. Pulmonary metastases: effect of CT section thickness on measurement——initial experience[J]. Radiology, 2005, 234(3):934-939.
- [18] Osborn JF. Chi 2 tests: how useful are they in the analysis of medical research data? [J]. Ann Ig, 1989, 1(3-4):417-432.
- [19] Neely JG, Hartman JM, Forsen JW, et al. Tutorials in clinical research: VII. Understanding comparative statistics (contrast)——part B: application of T-test, Mann-Whitney U, and chi-square [J]. Laryngoscope, 2003, 113(10):1719-1725.
- [20] Liao YY, Lee TS, Lin YM. A Fisher exact test will be more proper[J]. Radiology, 2006, 239(1):300-301.
- [21] Mullerad M, Hricak H, Wang L, et al. Prostate cancer: detection of extracapsular extension by genitourinary and general body radiologists at MR imaging[J]. Radiology, 2004, 232(1):140-146.
- [22] McElduff F, Cortina-Borja M, Chan SK, et al. When t-tests or Wilcoxon-Mann-Whitney tests won't do[J]. Adv Physiol Educ, 2010, 34(3):128-133.
- [23] Wilcoxon F. Probability tables for individual comparisons by ranking methods[J]. Biometrics, 1947, 3(3):119-122.
- [24] Wang L, Mazaheri Y, Zhang J, et al. Assessment of biologic aggressiveness of prostate cancer: correlation of MR signal intensity with Gleason grade after radical prostatectomy[J]. Radiology, 2008, 246(1):168-176.
- [25] Zhang J, Tehrani YM, Wang L, et al. Renal masses: characterization with diffusion-weighted MR imaging——a preliminary experience[J]. Radiology, 2008, 247(2):458-464.
- [26] Bishara AJ, Hittner JB. Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation and resampling approaches[J]. Psychol Methods, 2012, 17(3):399-417.
- [27] Zhang J, Kang SK, Wang L, et al. Distribution of renal tumor growth rates determined by using serial volumetric CT measurements[J]. Radiology, 2009, 250(1):137-144.
- [28] Zhou Y, Kierans A, Kenul D, et al. Mild traumatic brain injury: longitudinal regional brain volume changes[J]. Radiology, 2013, 267(3):880-890.
- [29] Huang W, LaBerge JM, Lu Y, et al. Research publications in vascular and interventional radiology: research topics, study designs and statistical methods[J]. J Vasc Interv Radiol, 2002, 13(3):247-255.
- [30] Kuss O. Global goodness-of-fit tests in logistic regression with sparse data[J]. Stat Med, 2002, 21(24):3789-3801.
- [31] Pulkstenis E, Robinson TJ. Two goodness-of-fit tests for logistic regression models with continuous covariates [J]. Stat Med, 2002, 21(1):79-93.
- [32] Hall CB, Zeger SL, Bandeen-Roche KJ. Adjusted variable plots for Cox's proportional hazards regression model[J]. Lifetime Data Anal, 1996, 2(1):73-90.
- [33] Wang L, Hricak H, Kattan MW, et al. Prediction of seminal vesicle invasion in prostate cancer: incremental value of adding endorectal MR imaging to the Kattan nomogram[J]. Radiology, 2007, 242(1):182-188.
- [34] Fuchsjäger MH, Shukla-Dave A, Hricak H, et al. Magnetic resonance imaging in the prediction of biochemical recurrence of prostate cancer after radical prostatectomy[J]. BJU Int, 2009, 104(3):315-320.

(收稿日期:2013-08-13 修回日期:2013-08-26)