

论著·临床研究 doi:10.3969/j.issn.1671-8348.2016.06.017

4 种评分系统对胃癌根治术手术风险的预测价值

晏佳¹, 王亚旭^{2△}

(1. 重庆医科大学附属大学城医院消化中心, 重庆 401331; 2. 重庆医科大学附属第二医院胃肠外科, 重庆 400010)

[摘要] **目的** 比较生理学和手术严重程度评分系统(POSSUM)、p-POSSUM、o-POSSUM、急性生理学及慢性健康状况评分 II (APACHE II) 4 种评分系统在对接受胃癌根治术后患者的并发症率和病死率中的预测价值。**方法** 通过使用 POSSUM、p-POSSUM、o-POSSUM 及 APACHE II 4 种评分系统对 306 例术前或术中病理诊断已明确并接受胃癌根治术的患者进行评估。对评分系统的预测能力将使用受试者特征面积(ROC)s、Hosmer-Lemeshow χ^2 拟合优度检验, t 检验和 χ^2 检验进行评估。**结果** 并发症率总体为 33.99%, POSSUM 系统的预测并发症率平均值为 36.61% [总体实际值/预测值(O/E) = 0.93, $P > 0.05$, AUCs = 0.79, Hosmer-Lemeshow 拟合优度检验 $P > 0.05$]。病死率总体为 2.94%, p-POSSUM 的预测病死率平均值为 3.22% (总体 O/E = 0.91, $P > 0.05$, AUCs = 0.76, Hosmer-Lemeshow 拟合优度检验 $P > 0.05$) 和 o-POSSUM 的预测病死率平均值 2.32% (总体 O/E = 1.26, $P > 0.05$, AUCs = 0.88, Hosmer-Lemeshow 拟合优度检验 $P > 0.05$) 相当, 与 POSSUM 的平均病死率预测值 8.51% (总体 O/E = 0.34, $P < 0.05$, AUCs = 0.63, Hosmer-Lemeshow 拟合优度检验 $P > 0.05$) 和 APACHE II 的平均预测值 10.31% (总体 O/E = 0.28, $P < 0.05$, AUCs = 0.63, Hosmer-Lemeshow 拟合优度检验 $P < 0.05$) 相比, 在预测病死率中有较高的价值。**结论** 预测胃癌根治术后并发症率时 POSSUM 评分系统的预测价值较高; 预测病死率时 p-POSSUM 和 o-POSSUM 的预测价值较高。

[关键词] 胃癌切除术; 评分系统; 受试者工作特征曲线; 受试者工作特征曲线下面积**[中图分类号]** R730.56 **[文献标识码]** A **[文章编号]** 1671-8348(2016)06-0775-04

Predictive value of four kinds of scoring system to operative risk of gastric cancer radical operation

Yan Jia¹, Wang Yaxu^{2△}

(1. Digestion Center, Affiliated University-Town Hospital, Chongqing Medical University, Chongqing 401331 China; 2. Department of Gastrointestinal Surgery, Second Affiliated Hospital, Chongqing Medical University, Chongqing 400010, China)

[Abstract] **Objective** To compare the values among the four scoring systems of physiological and operative severity score for enumeration of mortality and morbidity(POSSUM), portsmouth-POSSUM(p-POSSUM), oesophagogastric-POSSUM(o-POSSUM) and acute physiology and chronic health evaluation II (APACHE II) in predicting the postoperative complications rate and mortality rate in the patients undergoing gastric cancer radical operation. **Methods** A total of 306 adult cases of gastric cancer diagnosed by preoperative electronic gastroscopy or intraoperatively pathologic examination and undergone radical gastrectomy were evaluated by using the POSSUM, p-POSSUM, o-POSSUM, APACHE II scoring systems. The area of under curve in receiver operating characteristic(ROC)s curve, Hosmer-Lemeshow χ^2 test of goodness of fit, t -test and χ^2 -test analyses were used to evaluate the predictive ability of these scoring systems. **Results** The overall complication rate was 33.99%, the mean predictive complication rates by POSSUM was 36.61% (O/E = 0.93, $P > 0.05$, AUCs = 0.79, Hosmer-Lemeshow test $P > 0.05$). The overall mortality rate was 2.94%, the mean predictive mortality rate of p-POSSUM was 3.22% (O/E = 0.91, $P > 0.05$; AUCs = 0.76, Hosmer-Lemeshow test $P > 0.05$) and which of o-POSSUM was 2.32% (O/E = 1.26, $P > 0.05$, AUCs = 0.88, Hosmer-Lemeshow test $P > 0.05$), showing that their predictive mortality rates were similar and these two scoring systems had higher value for predicting mortality rate compared with 8.51% in POSSUM (O/E = 0.34, $P < 0.05$, AUCs = 0.63, Hosmer-Lemeshow test $P > 0.05$) and 10.31% in APACHE II (O/E = 0.28, $P < 0.05$, AUCs = 0.63, Hosmer-Lemeshow test $P < 0.05$). **Conclusion** The POSSUM scoring system has a higher value for predicting the complication rate following radical surgery of gastric cancer. And both the p-POSSUM and o-POSSUM scoring systems have a higher value in predicting postoperative mortality than the other two scoring systems.

[Key words] radical gastrectomy; scoring system; receiver operating characteristic curve; area under the curve

全世界范围内最常见恶性肿瘤中, 胃癌居第 4 位^[1]。中国胃癌发病率存在明显的地区和城乡差别^[2]。目前, 胃癌的最有效治疗方法仍为手术治疗。在临床诊疗活动中对手术风险的预测更显重要。本研究将应用生理学和手术严重程度评分系统(physiological and operative severity score for enumeration of mortality and morbidity, POSSUM)、portsmouth-POSSUM (p-POSSUM)、oesophagogastric-POSSUM(o-POSSUM)、急性生理学及慢性健康状况评分 II (acute physiology and chronic

health evaluation II, APACHE II) 4 种评分系统对接受胃癌根治术的 306 例患者进行分析。因上述 4 种评分系统中仅有 POSSUM 评分系统能对术后并发症率进行评估, 故此研究仅检测 POSSUM 评分系统的术后并发症率的准确性; 在预测胃癌根治术后的病死率中作者将比较以上 4 种评分系统的预测价值。

1 资料与方法

1.1 一般资料 对 306 例在重庆医科大学附属第二医院胃肠

外科接受胃癌根治术(D2 术式)患者的资料进行回顾性分析,以上患者的主刀医师和一助均为同一组医师,并在术前或术中行病检确诊为恶性肿瘤。将入院 24 h 内因穿孔、出血、梗阻等原因接受手术的患者分为急诊组,其他为择期组。两组患者手术方式等基线资料比较,见表 1。

表 1 两组患者手术方式等基线资料比较

项目	择期组(n=255)	急诊组(n=51)	合计(n=306)
男/女(n)	152/103	30/21	182/124
BMI($\bar{x}\pm s$,kg/m ²)	25.11±2.02	24.81±2.21	25.12±2.01
年龄($\bar{x}\pm s$,岁)	61.20±13.20	64.02±12.91	61.72±13.11
手术方式[n(%)]			
全胃切除(D2)	50(19.61)	6(11.76)	56(18.30)
近端胃大部切除(D2)	110(43.14)	20(39.22)	130(42.48)
远端胃大部切除(D2)	95(37.25)	25(49.02)	120(39.22)

1.2 方法

1.2.1 围术期处理 术前对所有患者的生理学指标进行纠正、控制血糖、血压,并留置胃管及尿管。术后所有患者均到 ICU 监护 24 h 或更长时间,并给予胃肠外营养、预防感染和控制血糖、血压等治疗,待病情平稳后转回病房。

1.2.2 评分系统 POSSUM^[3]评分系统由 12 个生理学指标及 6 个手术学指标组成,可预测手术的术后并发症率及病死率;p-POSSUM^[4]和专为胃食管癌设计的 o-POSSUM 在 POSSUM 的基础上进行改良,纠正 POSSUM 评分系统高估术后病死率的现象^[4-5],p-POSSUM 通过不同的计算公式计算病死率^[5-6],o-POSSUM^[7]通过重新筛选计算指标和计算公式计算病死率;APACHE II^[8]评分系统包括 12 个急性生理评分、年龄评分及慢性健康评分,并有研究证明可用于预测术后病死率。本次研究的评估数据是术前最新的检查结果。

1.2.3 术后评估 术后 30 d 内出现的并发症定义为手术并发症。包括由手术切口引起但无全身系统反应的局部并发症和影响了全身系统或需再次手术的系统并发症(表 2);术后 30 d 内的死亡即为手术死亡。共有 9 例因肺部感染、消化道出血、败血症等导致多器官功能衰竭的死亡患者,总体病死率为 2.94%。

1.2.4 评估方法 评价评分方法效能的两个特异性指标包括分辨度(discrimination,DIS)和校准度(calibration,CAL)。DIS 是指某一评价系统将可能死亡的患者与可能存活的患者区分开来的能力。临床上通常用受试者工作特征曲线(ROC)下面积(AUCs)反映真阳性率和假阴性率之间的关系。AUCs 越大,表明该评分方法的 DIS 越大,AUCs ≥ 0.8 ,模型有非常好

的预测准确性,0.7 \leq AUCs < 0.8 ,有临床应用价值;CAL 是指某一危重症评价系统预计的院内病死率(PHM)与患者实际病死率之间的吻合程度,通常用 Hosmer-Lemeshow 拟合优度检验(good of fit test)进行评价,Hosmer-Lemeshow χ^2 拟合优度检验中 $P < 0.05$ 认为预测校准度不佳。如果一个病情评价系统的 DIS 和 CAL 均较高,说明其特异性越高。

1.3 统计学处理 采用 SPSS19.0 统计软件进行统计学处理,计量资料用 t 检验,计数资料用 χ^2 检验,以 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 POSSUM 预测并发症率情况 POSSUM 的总体预测并发症率平均值为 36.61%,实际并发症率为 33.99%,实际值/预测值(O/E)=0.93($P > 0.05$),在各个亚组中并发症率的 O/E 差异无统计学意义($P > 0.05$),见表 3。POSSUM 评分系统预测并发症率的 AUCs=0.79,Hosmer-Lemeshow 拟合优度检验, $\chi^2 = 15.61$,自由度 7, $P = 0.46$ 。POSSUM 评分系统预测并发症率的 ROC,见图 1。

表 2 术后并发症分布情况(n)

并发症	择期	急诊	总计
局部并发症			
切口出血	3	1	4
切口感染	14	6	20
切口液化	13	7	20
系统并发症			
呼吸衰竭	4	3	7
深部出血	4	2	6
肺部感染	15	8	23
尿路感染	10	7	17
败血症	3	1	4
不明原因发热	2	0	2
心脏衰竭	3	1	4
肾功能损害	3	0	3
低血压	3	0	3
吻合口漏	5	0	5
肠梗阻	2	0	2
总体并发症	84	36	120
再手术	6	0	6

表 3 POSSUM 预测不同亚组的并发症发生情况[n(%)]

项目	年龄		TNM 分级		手术急缓		总计 (n=306)
	老年组 (≥ 60 岁,n=135)	中青年组 (< 60 岁,n=171)	I/II (n=185)	III/IV (n=121)	择期 (n=255)	急诊 (n=51)	
实际并发症	64(47.41)	40(23.39)	46(24.86)	58(47.93)	80(31.37)	24(47.06)	104(33.99)
预测并发症	60(44.44)	52(30.41)	63(34.05)	49(40.50)	83(32.55)	29(56.86)	112(36.61)
O/E	1.06	0.77	0.73	1.19	0.96	0.83	0.93
P	0.63	0.14	0.05	0.24	0.78	0.32	0.50

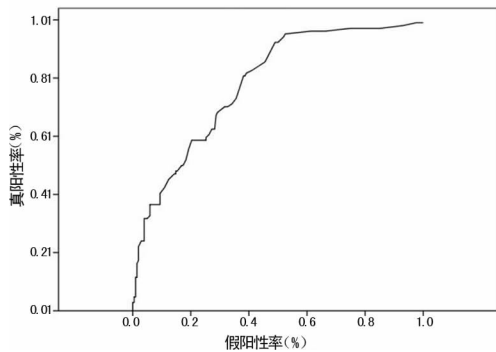


图 1 POSSUM 评分系统预测并发症率的 ROC

2.2 4 种评分系统预测病死率情况 实际病死率为 2.94% ($n=9$), POSSUM、p-POSSUM、o-POSSUM、APACHE II 评分系统预测病死率分别为 8.51%、3.22%、2.32%、10.31%。POSSUM 和 APACHE II 均过高预测各个亚组中的术后病死率($P<0.05$), p-POSSUM、o-POSSUM 未过高预测术后病死率($P>0.05$), 见表 4。p-POSSUM 和 O-POSSUM 的 AUCs 分别为 0.76、0.88, 预测准确性较好, POSSUM 和 APACHE II 的 AUCs 分别为 0.63、0.63, 有临床应用价值; Hosmer-Lemeshow 拟合优度检验校准时, APACHE II 预测校准度不佳($P<0.05$), 其余 3 项评分系统预测校准度较高($P>0.05$), 见表 5、6、图 2。

表 4 不同亚组中的实际病死率与预测病死率[n(%)]

组别	年龄		TNM 分级		手术急缓		总计 ($n=306$)
	老年组 (≥ 60 岁, $n=135$)	中青年组 (<60 岁, $n=171$)	I / II ($n=185$)	III / IV ($n=121$)	择期 ($n=255$)	急诊 ($n=51$)	
实际观察数	8(5.93)	1(0.58)	3(1.62)	6(4.96)	7(2.75)	2(3.92)	9(2.94)
POSSUM 预测数	16(11.92)	10(5.79)	11(5.88)	15(12.79)	18(7.01)	8(15.89)	26(8.51)
O/E	0.50	0.10	0.27	0.39	0.39	0.25	0.34
P	0.09	0.00	0.03	0.04	0.02	0.04	0.00
p-POSSUM 预测数	9(6.92)	1(0.43)	5(2.53)	5(4.42)	6(2.51)	4(7.24)	10(3.22)
O/E	0.86	1.50	0.64	1.14	1.08	0.54	0.91
P	0.80	1.00	0.48	0.76	0.78	0.40	0.82
o-POSSUM 预测数	6(4.21)	1(0.82)	3(1.41)	4(3.62)	5(1.84)	2(4.41)	7(2.32)
O/E	1.40	0.75	1.14	1.39	1.50	0.89	1.26
P	0.58	1.00	1.00	0.52	0.56	1.00	0.61
APACHE II 预测数	18(13.12)	14(8.41)	12(6.33)	20(16.81)	22(8.81)	10(19.01)	32(10.31)
O/E	0.45	0.07	0.25	0.30	0.31	0.21	0.28
P	0.04	0.00	0.02	0.00	0.00	0.01	0.00

表 5 各评分系统预测病死率的 AUCs

检测结果变量	曲线下面积	标准误	95%CI
POSSUM	0.63	0.10	0.44~0.71
p-POSSUM	0.76	0.07	0.73~0.81
o-POSSUM	0.88	0.04	0.81~0.95
APACHE II	0.63	0.10	0.44~0.72

表 6 各评分系统预测病死率的拟合优度检验

检测结果变量	χ^2	自由度	P
POSSUM	6.77	7	0.45
p-POSSUM	4.49	8	0.81
o-POSSUM	14.68	5	0.51
APACHE II	6.24	7	0.03

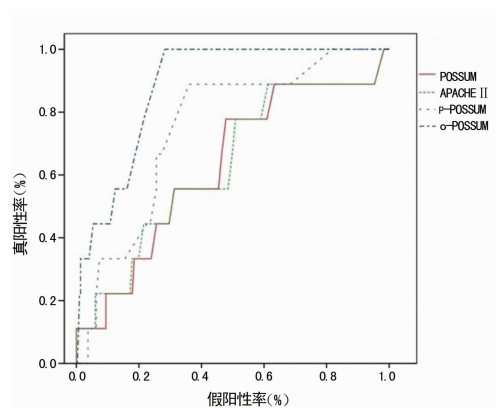


图 2 4 种评分系统预测病死率的 ROC

3 讨 论

虽然新型药物的研发和手术方式及设备的改进不断减小了麻醉和手术过程中出现危险性的风险。但患者自身的某种危险因素不会随着技术的进步迎来太大的改变, 如患有慢性疾病患者的手术风险。所以选择有预测价值的评分系统将在降低围术期的并发症率及病死率中起到明显的指导作用。

在临床使用过程中, POSSUM 评分系统的并发症率及病死率的预测价值有较好的表现。分析术后并发症率时, 本文发现 POSSUM 评分系统总体的预测值与实际值相似($O/E=0.93, P>0.05$)、 $AUCs=0.79$ 、Hosmer-Lemeshow 拟合优度检验 $P>0.05$, 以上统计学指标均提示较好预测价值, 同时在年龄、TNM 分级和手术急缓的各个亚组中也有较高的预测价值。本文结果与文献[9-10]利用 POSSUM 评分系统预测胃癌根治术后的并发症率的结果相似(O/E 值分别为 0.99、0.83),

但与德国和英国报道的并发症率(O/E 值分别为 0.52、0.40)结果相反^[11-12]。作者认为地区差异可能导致了上述结果的偏差,在预测亚洲患者的术后并发症率中 POSSUM 可能有较高的预测价值。

在术后病死率的预测中,POSSUM 评分系统总体过高预测了病死率发生(O/E=0.34, $P<0.05$)。在中青年组中,术后病死率预测值过高的情况(O/E=0.10, $P<0.01$)与国外的结果相似^[9,11-13],提示该系统在预测低危人群时高估了病死率的发生风险。在 TNM 分级和手术急缓分组中,POSSUM 预测低危人群和高危人群术后病死率时均出现过高估计现象(AUCs=0.63, Hosmer-Lemeshow 拟合优度检验 $P>0.05$),以上指标提示该系统的分辨率较差,校准度尚可。出现以上差异,作者认为可能原因是该系统的设计是用于预测接受普通手术后出现相关风险的危险性。但在胃癌根治术中,该系统中的手术严重程度分值过高,而该分值与病死率的预测值呈正相关;同时该评分系统的部分项目,如胸片,心电图及 Glasgow 评分等,并不能较好反映接受胃癌根治术时可能出现的危险性。上述过高或过低的分值组成均可能造成 POSSUM 评分系统在预测术后病死率时的偏差。在预测各亚组和总体术后病死率中,p-POSSUM 和 o-POSSUM 均有较高预测价值,从表 4 和表 5 中可得知 p-POSSUM 和 o-POSSUM 的预测分辨率较好(AUCs 分别为 0.76, 0.88, $P>0.05$),且 o-POSSUM 分辨率高于 p-POSSUM,二者渐进 95% CI 无重叠,提示差异有统计学意义($P<0.05$);Hosmer-Lemeshow 拟合优度检验以上两种评分系统均提示 $P>0.05$,表明二者预测准确度较好。Goemans 等^[13]在对接受胃癌根治术患者使用 P-POSSUM 预测术后病死率时出现过高预测上述风险的情况(O/E=0.70, $P<0.05$, AUCs=0.70),而 o-POSSUM 却有较好的预测价值(O/E=0.90, $P>0.05$, AUCs=0.880)。以上差异作者考虑原因如下:(1)本次研究中的预测率与实际率差异虽然较大,但受样本量较小的限制,相应人数之间在进行统计学差异分析时并无明显差异。(2)本次评分系统中采用的数据均来源于最靠近手术时间的经过术前纠正异常结果后的数据,与既往入院即采集数据的要求不同,术前经过对生化指标、血红蛋白、白细胞、血糖等进行纠正后,评分系统中的相应加权分值有降低,故使得本次研究的术后预测病死率与国外相比更接近实际值,同时也解释在急诊组中术后病死率的预测值均较择期组大的现象。本次研究中 APACHE II 在预测接受胃癌根治术患者的术后病死率时表现较差,在任何亚组和总体分析中,其 O/E 值均较低,均 $P<0.05$, AUCs=0.63, Hosmer-Lemeshow 拟合优度检验 $P<0.05$,以上指标综合提示该评分系统的预测分辨率和预测准确度均较低,过高估计了本次接受胃癌根治术患者的术后病死率。以上现象的产生可能系该系统组成项目中未包括可能增加手术风险的指标,如心血管方面的急性缺血性心电图改变,严重哮喘的存在,患者当前的营养状况和有无急性心肌梗死的病史^[14]。同时,使用 APACHE II 评分系统要求使用数据应为初入 ICU 时 24 h 内最差的结果,该要求也同样增加预测手术风险的偏差。

在本次调查接受胃癌根治术的患者中,POSSUM 评分系统在术后并发症率的预测中有较高价值;在预测术后病死率中,p-POSSUM 和 o-POSSUM 有较高预测价值,其中 o-POSSUM 较 p-POSSUM 有更高的预测价值,POSSUM 和 APACHE II 预测价值较差。但本次调查样本量较小,需要更大的样本量来评价 POSSUM 和 APACHE II 评分系统在术后病死率预测中的价值。

参考文献

- [1] Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world[J]. *J Clin Oncol*, 2006, 24(14): 2137-2150.
- [2] 吴春晓, 郑莹, 鲍萍萍, 等. 上海市胃癌发病流行现况与时间趋势分析[J]. *外科理论与实践*, 2008, 13(1): 24-29.
- [3] Kodama A, Narita H, Kobayashi M, et al. Usefulness of POSSUM physiological score for the estimation of morbidity and mortality risk after elective abdominal aortic aneurysm repair in Japan[J]. *Circ J*, 2011, 75(3): 550-556.
- [4] Brooks MJ, Sutton R, Sarin S. Comparison of Surgical Risk Score, POSSUM and p-POSSUM in higher-risk surgical patients[J]. *Br J Surg*, 2005, 92(10): 1288-1292.
- [5] Hellmann S, Schafmayer C, Hinz S, et al. Evaluation of the POSSUM score in surgical treatment of cholangiocarcinoma[J]. *Hepatogastroenterology*, 2010, 57(5/6): 403-408.
- [6] Lai F, Kwan TL, Yuen WC, et al. Evaluation of various POSSUM models for predicting mortality in patients undergoing elective oesophagectomy for carcinoma[J]. *Br J Surg*, 2007, 94(9): 1172-1178.
- [7] Lagarde SM, Maris AK, de Castro SM, et al. Evaluation of o-POSSUM in predicting in-hospital mortality after resection for oesophageal cancer[J]. *Br J Surg*, 2007, 94(12): 1521-1526.
- [8] Adesunkanmi AR, Badmus TA, Fadiora FO, et al. Generalized peritonitis secondary to typhoid ileal perforation: Assessment of severity using modified APACHE II score[J]. *Indian J Surg*, 2005, 67(1): 29-33.
- [9] Otsuka Y, Kunisaki C, Ono H, et al. Modified POSSUM to predict postoperative morbidity following gastrectomy[J]. *Hepatogastroenterology*, 2007, 54(76): 1142-1145.
- [10] Sah BK, Min CM, Yan WX, et al. Risk adjusted auditing of postoperative complications in gastric cancer patients by POSSUM[J]. *Int J Surg*, 2008, 6(4): 311-316.
- [11] Bollschweiler E, Lubke T, Monig SP, et al. Evaluation of POSSUM scoring system in patients with gastric cancer undergoing D2-gastrectomy[J]. *BMC Surg*, 2005, 5(1): 8-12.
- [12] Lamb P, Sivashanmugam T, White M, et al. Gastric cancer surgery—a balance of risk and radicality[J]. *Ann R Coll Surg Engl*, 2008, 90(3): 235-242.
- [13] Gomes A, Rocha R, Marinho R, et al. Colorectal surgical mortality and morbidity in elderly patients: comparison of POSSUM, P-POSSUM, CR-POSSUM and CR-BHOM[J]. *Int J Colorectal Dis*, 2015, 30(2): 173-179.
- [14] Mayo SC, Pawlik TM. Current management of colorectal hepatic metastasis[J]. *Expert Rev Gastroenterol Hepatol*, 2009, 3(2): 131-144.