

als.

【Key words】 Perihilar cholangiocarcinoma; Surgery; Composite endpoint of liver surgery; Clinical trial; Sample size

肝门部胆管癌(pCCA)占胆道恶性肿瘤的 50%~60%,发病率持续上升^[1]。随着影像学与分子诊断技术的进步,更多早期 pCCA 患者获得根治性手术机会,但肝切除术导致的胆管离断、残余肝体积不足及血管损伤等问题仍引发肝脏特异性并发症,包括胆漏、肝切除术后肝衰竭(PHLF)、肝切除术后出血(PHH)等,显著增加术后死亡率并影响远期生存^[2~6]。精准评估手术质量对改善预后至关重要。传统评估体系依赖单一结局,虽具临床价值,却难以全面反映手术整体质量。现代外科技术的多维进步进一步暴露单一指标的局限性^[7,8]。此外,验证手术创新的高级别临床研究常受限于样本量需求:单一结局发生率低,导致随机对照试验(RCT)需要数千例样本,这在肝胆外科实践中难以实现^[9]。为此,Pawlik 团队开发了肝脏手术复合终点(CELS),整合失血量 ≥ 2000 ml、胆漏、PHLF 和 PHH 四项指标^[10]。该指标在肝细胞癌等疾病中已验证可降低临床试验样本量,并有效预测手术死亡率。然而,pCCA 具有独特的解剖复杂性和更高的并发症风险,CELS 的适用性尚未验证^[11,12]。本研究评估 CELS 对 pCCA 根治性切除术的质量评价效能,旨在为建立 pCCA 手术标准化评估体系提供循证依据。

1 对象与方法

1.1 研究对象 连续纳入 2019 年 1 月至 2022 年 1 月在我院接受根治性目的切除术的 pCCA 患者。纳入标准:①经病理检查证实为 pCCA;②接受根治性目的切除术。排除标准:①姑息性手术(R2 切除);②失去围手术期数据和随访数据;③基线资料缺失 $>10\%$ 。此项研究获得医院伦理审查委员会的批准(ID:KY2021129)。

1.2 根治性目的切除术 pCCA 的可切除性评估基于残余肝脏体积、肿瘤与血管的解剖关系、肝脏功能、机体状态功能评估^[13]。根治性目的切除的术式包括尾状叶切除、肝切除术、区域淋巴结清扫、胆管切除与重建。若肿瘤侵犯大血管,则进行大血管切除与重建。肝切除的范围根据 Bismuth 分型进行评估。Bismuth I 型患者一般进行哑铃状切除;Bismuth II 型患者选择左、右半肝切除或中肝切除;Bismuth III 型和 IV 型患者行半肝切除或扩大半肝

切除。左/右半肝切除根据“新世界术语”由 H234/H5678 定义,扩大右半肝切除包括右半肝切除+部分 IV 段切除,扩大左半肝切除包括左半肝切除+部分肝段 V 和 VIII 切除,中肝叶切除为 IV、V 和 VIII 段切除,哑铃状切除为 IVB 段和部分 V 段切除。由于多中心性质,手术标准流程与器械不统一、手术水平的差异性无法避免。

1.3 变量收集与定义 收集患者的基线资料。临床特征包括年龄、性别、体质量指数(BMI)、美国麻醉医师协会(ASA)评分、肝硬化、术前减黄(PTCD)。实验室指标包括总胆红素(TB)、丙氨酸氨基转移酶(ALT)、国际标准化比值(INR)、白蛋白(ALB)、碳水化合物抗原 19-9(CA19-9)。手术相关变量包括手术方式(即右半肝切除、扩大右半肝切除、左半肝切除、扩大左半肝切除、中肝叶切除或哑铃状切除)、淋巴结清扫数量、微创手术及血管切除和重建的使用情况。病理学变量包括淋巴结受累情况、微血管侵犯、最大肿瘤大小、肿瘤分化程度、第八届美国癌症联合委员会(AJCC)分期和 Bismuth 分型。围手术期结局变量包括术中出血量、胆漏、PHLF、PHH、并发症、手术相关死亡。肝脏手术相关并发症基于国际肝脏手术研究组(ISGLS)的共识来定义,包括胆漏、PHLF、PHH。所有并发症的严重程度均根据 Clavien-Dindo 分级系统(I~V 级)定义^[14]。严重并发症定义为 Clavien-Dindo 分级 \geq III 级。

1.4 随访 接受根治性目的切除术后,患者在术后前 2 年内每 3~4 个月进行一次监测,之后每 6 个月监测一次。在随访评估期间,通过检测血清肿瘤标志物和影像学检查来监测肿瘤^[15]。总生存期(OS)定义为从手术日期至死亡日期或末次随访日期或死亡的时间间隔。

1.5 研究终点和预测终点 主要研究终点为手术相关死亡,定义为术后 90 天内因任何并发症导致的死亡。本研究预测终点是 CELS^[10],纳入了肝脏手术特异性并发症,包括:胆漏、PHLF、PHH、失血量 ≥ 2000 ml。失血量 ≥ 2000 ml 的临界值是参考 Oslo 分类确定的。如果患者经历失血量 ≥ 2000 ml、胆漏、PHLF、PHH 中的一种或多种,被定义为 CELS 阳性;如果患者没有经历失血量 ≥ 2000 ml、胆漏、PHLF、PHH 中的任何一种,则定义为 CELS 阴性。

1.6 统计学方法 Student's *t* 检验或 Mann-

【基金项目】 重庆市自然科学基金面上项目(编号: CSTB2022NSCQ-MSX0204)

【通讯作者】 刘智鹏

Whitney *U* 检验用于连续变量, Pearson 卡方检验用于分类变量。连续变量在适合的情况下以中位数 (Q_1, Q_3) 或均数 \pm 标准差表示, 分类变量以频数 (%) 表示。多因素 Logistic 回归用于评估临床相关预测变量与手术相关死亡之间的关联。使用敏感性、特异性和受试者操作特征 (ROC) 曲线以及曲线下面积 (AUC) 评估模型区分手术相关死亡患者的能力。使用 Kaplan-Meier 曲线和对数秩检验评估 CELS 对 OS 的影响。Logistic 回归模型对 CELS 预测手术相关死亡的校准进行了评估。使用 SPSS 软件 (2023 版) 计算在 RCT 中每个试验组所需的最少患者数量, 以此确定理论样本量, 假设显著性水平 (α) 为 0.05, 检验功效 (1-beta) 为 0.80 (即 beta 为 0.20), 并采用双侧检验。样本量的计算是基于假设单一组成部分或新开发的 CELS 的发生率相对降低 10%、20%、30%、40%、50% 或 60%。其余统计分析均使用 R 软件 (4.4.0 版 <http://www.r-project.org/>) 进行。所有检验均为双侧检验, $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 患者特征 共纳入 389 例 pCCA 接受根治性目的切除术的患者。手术方式包括右半肝切除 ($n = 9, 2.3\%$)、扩大右半肝切除 ($n = 64, 16.5\%$)、左半肝切除 ($n = 23, 5.9\%$)、扩大左半肝切除 ($n = 177, 45.5\%$)、中肝叶切除 ($n = 61, 15.7\%$) 以及哑铃状切除 ($n = 55, 14.1\%$)。所有患者的基线资料见表 1。术中出血量 600 ml (400, 950 ml), 其中, 27 (6.9%) 患者的失血量 ≥ 2000 ml。257 例 (66.1%) 患者发生并发症, 其中, 130 例 (33.4%) 患者为 Clavien-Dindo III、IV 级、V 级并发症 (严重并发症)。55 例 (14.1%) 患者发生胆漏, 24 例 (6.1%) 患者发生 PHLF, 47 例 (12.1%) 患者发生 PHH, 29 例 (7.5%) 患者发生了手术相关死亡。见表 2。

2.2 与 CELS 相关的多因素分析 本研究的 389 例患者中 124 例 (31.9%) 为 CELS 阳性。将 CELS 作为预测指标, 终点结局为手术相关死亡, 发现在本研究队列中 CELS 能够很好地预测手术相关死亡, 其 AUC 为 0.738 (95% CI: 0.644 ~ 0.832), 敏感性为 0.759 (95% CI: 0.561 ~ 0.890), 特异性为 0.717 (95% CI: 0.667 ~ 0.762)。与手术相关死亡相关的独立危险因素包括: 失血量 ≥ 2000 ml、PHLF 和 PHH ($P < 0.05$)。胆漏不是与手术相关死亡的独立危险因素 ($P = 0.956$)。见表 3。

2.3 CELS 与长期结局 在 389 例患者中, 中位随访时间为 23.0 个月 [(10.9, 47.3) 月]。CELS 阳性的患者中位 OS 为 21.0 个月 (95% CI: 13.7 ~ 28.3

月), CELS 阴性患者的中位 OS 为 47.2 个月 (95% CI: 33.0 ~ 61.4 月)。CELS 阳性患者的 5 年 OS 率 35.3% 明显低于 CELS 阴性组患者 44.2% ($\chi^2 = 12.322, P < 0.001$)。见图 1。

表 1 肝门部胆管癌患者的基本特征 [n (%)]

特征	例数 ($n = 389$)
年龄 > 60 岁	188 (48.3)
男性	225 (57.8)
BMI ≥ 25.0 kg/m ²	56 (14.4)
ASA 评分 > II 级	22 (5.7)
肝硬化	8 (2.1)
PTCD	151 (38.8)
TB > 17.1 μ mol/L	359 (92.3)
ALT > 40 U/L	328 (84.3)
INR > 1.15	41 (10.5)
ALB < 35 g/L	99 (25.4)
CA19-9 > 37 U/L	326 (83.8)
手术方式	
右半肝切除	9 (2.3)
扩大右半肝切除	64 (16.5)
左半肝切除	23 (5.9)
扩大左半肝切除	177 (45.5)
中肝叶切除	61 (15.7)
哑铃状切除	55 (14.1)
血管切除和重建	48 (12.3)
微创手术	53 (13.6)
淋巴结侵犯	156 (40.1)
淋巴结清扫的数量	4 (2, 7)
微血管侵犯	54 (13.9)
肿瘤大小 > 3 cm	119 (30.6)
肿瘤低分化	66 (17.0)
第八期 AJCC 分级	
I 级	60 (15.4)
II 级	106 (27.2)
III 级	180 (46.3)
IV 级	43 (11.1)
Bismuth 分级	
I 级	58 (14.9)
II 级	57 (14.7)
III 级	211 (54.2)
IV 级	63 (16.2)

表 2 肝门部胆管癌患者的围手术期结局

结局	例数 (n = 389)
失血量 (ml)	600 (400, 950)
≥2000 ml [n(%)]	27 (6.9)
胆漏 [n(%)]	55 (14.1)
PHLF [n(%)]	24 (6.1)
PHH [n(%)]	47 (12.1)
并发症 [n(%)]	257 (66.1)
I 级	64 (16.5)
II 级	63 (16.2)
III 级	76 (19.5)
IV 级	25 (6.4)
V 级	29 (7.5)
严重并发症 [n(%)]	130 (33.4)
手术相关死亡 [n(%)]	29 (7.5)
CELS [n(%)]	124 (31.9)

2.4 样本量评估 根据并发症发生率计算临床试验的样本量要求,以实现不同水平下最小临床样本需求的降低。假设相对风险降低 10%,以 CELS 作为主要终点,所需的总样本量为 1158 例,而单独使用失血量 ≥ 2000 ml 作为终点需要 6974 例,单独使

用胆漏需要 3184 例,单独使用 PHLF 需要 7948 例,单独使用 PHH 需要 3786 例。假设相对风险降低 50%,所需的样本量分别为:使用 CELS 需要 60 例,单独使用失血量 ≥ 2000 ml 需要 314 例,单独使用胆漏需 148 例,单独使用 PHLF 需要 356 例,单独使用 PHH 需要 174 例。假设相对风险降低 10%、20%、30%、40%、50% 和 60%,本研究计算了不同主要终点所需的样本量。见表 4。

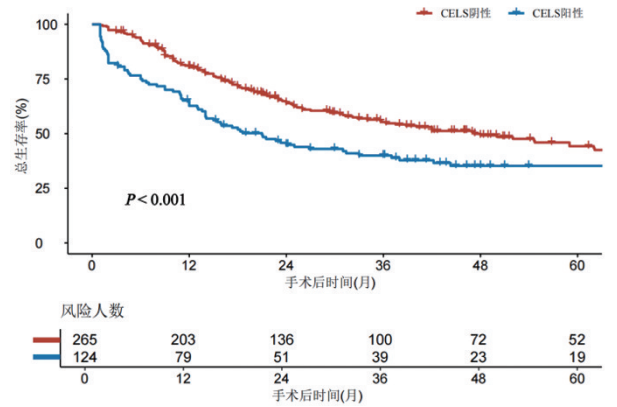


图 1 肝脏手术复合终点阳性与肝脏手术复合终点阴性患者的总生存率

表 3 与手术相关死亡的多因素 Logistic 回归分析

变量	β	SE	Wald χ^2	P	OR	95% CI
失血量 ≥ 2000 ml	1.163	0.539	2.160	0.031	3.201	1.114 ~ 9.198
胆漏	-0.031	0.559	-0.056	0.956	0.969	0.324 ~ 2.901
PHLF	4.269	0.537	7.597	< 0.001	71.440	24.961 ~ 204.468
PHH	1.880	0.417	4.510	< 0.001	6.555	2.895 ~ 14.839

表 4 基于各主要结局参数所假定的相对风险降低幅度进行样本量计算 (n)

相对风险降低	10%	20%	30%	40%	50%	60%
失血量 ≥ 2000 ml	6974	1804	826	478	314	222
胆漏	3184	832	384	224	148	106
PHLF	7948	2504	940	542	356	252
PHH	3786	986	454	264	174	124
CELS	1158	312	148	88	60	44

3 讨论

3.1 CELS 的开发 既往对于肝切除术,通常通过围手术期或肝脏特定的单一并发症直接评估手术质量^[16]。但是传统的单一结局提供的数据较为孤立,且较低的发生率使分析更具有偶然性,这也导致了它们通常无法全面地评估手术技术^[17]。此外,若基于单一结局通过临床试验来评估手术质量,则需要较大的样本量,从而限制了这些单一结局的实用性。为解决这些临床限制,Pawlik 团队开发和验证了一种新型复合结局,即 CELS,其不再以单一结局为研究终点,而将关键的肝脏特异性并发症整合在一起,结合失血量 ≥ 2000 ml、胆漏、PHLF、PHH

四个结局指标^[10]。若上述 4 个肝脏特异性并发症任意出现一个,则认为 CELS 阳性,若这 4 个肝脏特异性并发症均未出现,则认为 CELS 阴性。Pawlik 团队采用 CELS 评估了肝细胞肝癌、肝内胆管癌和结直肠癌肝转移患者行肝切除术的整体质量,证实 CELS 能够合理地预测手术相关死亡,并证明了 CELS 阳性患者的住院时间更长,最后明确了 CELS 相较于单一结局指标能够降低临床试验样本的需求量。pCCA 作为一种常见的毗邻肝门的肿瘤,根治性目的切除需要进行肝切除术^[18]。而在 pCCA 根治性目的切除术的质量评估,同样面临单一结局评估的局限性^[19,20]。因此,需要新的综合性结局来

弥补单一结局的不足。

3.2 CELS 的验证 本研究基于多中心数据库,明确 31.9% 的 pCCA 根治性目的切除术患者能够实现 CELS,验证了 CELS 对其手术相关死亡有较好的预测性能。相较于传统单一结局评估体系,CELS 的应用可大幅缩减临床试验所需样本量,从而为优化临床试验设计提供了新的方法学支持。具体而言,与传统单一结局相比,CELS 指标在评估手术结果方面具有多项优势。首先,CELS 能够精准预测患者的手术相关死亡(AUC:0.738),校准曲线也显示预测的手术相关死亡和实际的手术相关死亡具有良好的一致性。提示 CELS 对围手术期质量能够有理想的预测效果。第二,我们同时对 CELS 阳性和阴性患者的生存进行比较,发现 CELS 阳性患者的 5 年 OS 率明显低于 CELS 阴性组患者。提示 CELS 能够对远期生存也具有理想的预测效果。上述两个结果综合说明 CELS 能够为 pCCA 根治性目的切除术过程提供全面、连贯(从围手术期到远期)的质量评估。我们认为利用 CELS 作为手术质量评估的终点能够为患者制定更符合其收益的治疗方案。第三,使用 CELS 可以提高临床试验的统计功效。相比于传统的单一结局,CELS 可以减少 58.5% ~ 85.4% 所需的样本量。样本量的减少可以降低研究的成本并提高研究的效率,研究人员可以从小型队列中得到具有价值的结果,使得大型的前瞻性试验等的进行更具有实践性。

3.3 局限性 第一,回顾性分析无法避免相关的选择偏倚。这是回顾性验证研究中普遍存在的局限性,本研究利用多中心数据库的连续患者,尽可能纠正这种回顾性偏倚。第二,CELS 的开发基于西方人群,对于东方人群的适用性存在疑问。本研究基于东方人群进行验证,证实基于西方人群开发的 CELS 同样适用于东方人群。

综上,CELS 在 pCCA 根治性目的切除术中作为一种综合性结局能够很好地评估肝脏手术质量,能够较好地预测围手术期死亡。在实施临床试验的过程中,对比单一结局,CELS 进行差异性检验所需要的样本量更少,能够适用于小样本的外科临床试验。

【参考文献】

- [1] 中国研究型医院学会肝胆胰外科专业委员会,《中华消化外科杂志》编辑委员会. 肝门部胆管癌诊断和治疗指南(2025 版). 中华消化外科杂志,2025,24(1):1-20.
- [2] Liu ZP, Chen WY, Zhang YQ, et al. Postoperative morbidity adversely impacts oncological prognosis after curative resection for hilar cholangiocarcinoma[J]. World J Gastroenterol, 2022, 28(9): 948-960.
- [3] 邓志霖,裴煜,黄四洲. MCM 家族在 DNA 复制及肿瘤发生发展

- 中的研究进展[J]. 成都医学院学报,2024,19(3):534-537.
- [4] 金杯,张晔昱,潘景轩. 肝脏微环境细胞在肿瘤肝定植转移过程的作用及意义[J]. 四川大学学报(医学版),2023,54(3): 469-474.
- [5] van Keulen AM, Buettner S, Erdmann JJ, et al. Multivariable prediction model for both 90-day mortality and long-term survival for individual patients with perihilar cholangiocarcinoma: does the predicted survival justify the surgical risk[J]. Br J Surg, 2023, 10(5): 599-605.
- [6] Keulen AV, Gaspersz MP, van Vugt J, et al. Success, complication, and mortality rates of initial biliary drainage in patients with unresectable perihilar cholangiocarcinoma[J]. Surgery, 2022, 172(6):1606-1613.
- [7] Nickel F, Kuemmerli C, Müller PC, et al. The PANcreatic Surgery Composite Endpoint (PACE): Development and Validation of a Clinically Relevant Endpoint Requiring Lower Sample Sizes[J]. Ann Surg, 2025,281(3):496-500.
- [8] 何朗,李帅. 恶性肿瘤脑转移诊疗进展专家述评[J]. 西部医学,2024,36(9):1249-1255.
- [9] 王瑞平. 随机对照临床试验设计中的样本量估算方法[J]. 上海医药,2023,44(1):48-52.
- [10] Kawashima J, Akabane M, Endo Y, et al. A Composite Endpoint of Liver Surgery (CELS): Development and Validation of a Clinically Relevant Endpoint Requiring a Smaller Sample Size[J]. Ann Surg Oncol, 2025,32(5):3505 ~ 3515.
- [11] Jena SS, Mehta NN, Nundy S. Surgical management of hilar cholangiocarcinoma: Controversies and recommendations[J]. Korean journal of hepato-biliary-pancreatic surgery, 2023, 27(3):14.
- [12] 倪传斗,宋春峰,杨明星,等. 肝门部胆管癌外科治疗的预后影响因素分析[J]. 中华消化外科杂志,2019,18(1):83 ~ 90.
- [13] 中华医学会外科学分会胆道外科学组,解放军全军肝胆外科专业委员会. 肝门部胆管癌诊断和治疗指南(2013 版)[J]. 中华外科杂志,2013,51(10):865-871.
- [14] Clavien PA, Barkun J, de Oliveira ML, et al. The Clavien-Dindo classification of surgical complications: five-year experience[J]. Ann Surg, 2009,250(2):187-196.
- [15] 毛琼,杨翼飞,艾力西尔·艾拜都拉,等. 基于多学科诊断与治疗的肝门部胆管癌外科手术疗效和预后影响因素分析[J]. 中华消化外科杂志,2023,22(7):873 ~ 883.
- [16] Sucandy I, Marques HP, Lippert T, et al. Clinical Outcomes of Robotic Resection for Perihilar Cholangiocarcinoma: A First, Multi-center, Trans-Atlantic, Expert-Center, Collaborative Study[J]. Ann Surg Oncol, 2024,31(1):81-89.
- [17] 蒋弟,陈诚,刘智鹏等. 肝门部胆管癌根治性切除术后教科书式结局发生的预测因素分析[J]. 肝胆胰外科杂志,2023,35(5):263-267.
- [18] Maeda K, Kuriyama N, Ito T, et al. Safety and benefits of major hepatectomy with extrahepatic bile duct resection in older perihilar cholangiocarcinoma patients[J]. Langenbecks Arch Surg, 2022, 407(7):2861-2872.
- [19] 胡满琴,许丁伟,张燕,等. 机器人手术与开放手术治疗肝门部胆管癌疗效与安全性的荟萃分析[J]. 中华肝胆外科杂志, 2024,30(1):62-66.
- [20] Dixit S, Mahakalkar C, Kshirsagar S, et al. Efficacy and Safety of Robotic Surgery vs. Open Surgery for Hilar Cholangiocarcinoma: A Comprehensive Review[J]. Cureus, 2024,16(8):e66709.

(收稿日期:2025-04-09;修回日期:2025-05-12)

(本文编辑:彭羽)