

• 论 著 •

运用 Logistic 回归和 ROC 曲线综合评价 肿瘤标志物对肺癌的诊断价值*

陈佳¹, 向玲亚¹, 邵华侨¹, 白雪¹, 孙婧瑜¹, 陈柯煜², 何佳³, 芮东升³, 吴江东^{4△}
(石河子大学医学院: 1. 临床医学 2013 级; 2. 临床医学 2012 级; 3. 预防医学
系统计教研室; 4. 基础医学系人体解剖学教研室, 新疆石河子 832002)

摘要:目的 采用 Logistic 回归和受试者工作特征(ROC)曲线探讨癌胚抗原(CEA)、甲胎蛋白(AFP)、糖类抗原(CA)15-3、CA19-9、CA72-4 在肺癌中的诊断价值。**方法** 采用电化学发光免疫分析仪检测不同病理类型肺癌组、肺部良性疾病组及健康对照组血清肿瘤标志物(CEA、AFP、CA15-3、CA19-9、CA72-4)的水平, 通过 Logistic 回归建立回归模型, 用 ROC 曲线分析这些指标对肺癌诊断的意义。**结果** 肺癌组的 CEA、CA15-3、CA19-9、CA72-4 水平显著高于肺部良性疾病和健康对照组($P < 0.05$), 且 CEA 在肺癌的诊断中敏感性最高(45.16%)。建立回归模型 $Y = 1/[1 + \text{EXP}(1.982X_1 + 3.383X_2 + 2.373X_3 + 2.167X_4 - 1.679)]$, 其中 CA15-3 的 OR 最高, 所生成的新变量 Y 值在 ROC 曲线中的曲线下面积最大。**结论** AFP 对肺癌无临床诊断意义, 血清 CEA、CA15-3、CA19-9、CA72-4 联合检测对肺癌的诊断具有一定的价值, 综合运用 Logistic 回归和 ROC 曲线可以提高其肺癌诊断水平。

关键词: 肺癌; 肿瘤标志物; 癌胚抗原; Logistic 回归; 受试者工作特征曲线
DOI: 10.3969/j.issn.1672-9455.2017.05.003 **文献标志码:** A **文章编号:** 1672-9455(2017)05-0604-03

Comprehensive evaluation on diagnostic value of serum tumor markers in lung cancer by using Logistic regression and ROC curve*

CHEN Jia¹, XIANG Lingya¹, SHAO Huaqiao¹, BAI Xue¹, SUN Jingyu¹,
CHEN Keyu², HE Jia³, RUI Dongsheng³, WU Jiangdong^{4△}

(1. Grade 2013, Clinical Medicine; 2. Grade 2012, Clinical Medicine; 3. Teaching and Researching Section of Statistics, Faculty of Preventive Medicine; 4. Teaching and Researching Section of of Human Anatomy, Faculty of Basic Medicine, Medical College of Shihezi University, Shihezi, Xinjiang 832002, China)

Abstract: **Objective** To investigate the diagnostic value of serum CEA, AFP, CA15-3, CA19-9 and CA72-4 in diagnosing lung cancer by using the Logistic regression and ROC curve. **Methods** The levels of serum CEA, AFP, CA15-3, CA19-9 and CA72-4 in the different pathological types of lung cancer group, benign lungs diseases group and healthy control group were determined by adopting electrochemiluminescence immunoassay(ECLIA). The regression model was established by Logistic regression. The significance of these indicators in the diagnosis of lung cancer was analyzed by using the ROC curve. **Results** The levels of serum CEA, AFP, CA15-3, CA19-9 and CA72-4 in the lung cancer group were significantly higher than those in the benign lungs diseases group and healthy control group($P < 0.05$), moreover CEA had the highest sensitivity(45.16%) in the diagnosis of lung cancer. The regression model was constructed, $Y = 1/[1 + \text{EXP}(1.982X_1 + 3.383X_2 + 2.373X_3 + 2.167X_4 - 1.679)]$, in which OR of CA15-3 was highest, the area under the ROC curve of generated new variable Y was biggest. **Conclusion** Serum AFP had no significance in clinical diagnosis of lung cancer, the combined detection of CEA, CA15-3, CA19-9 and CA72-4 has a certain diagnostic value in lung cancer, and comprehensive use of Logistic regression and ROC curve can improve the diagnosis level of lung cancer.

Key words: lung cancer; tumor marker; carcinoma embryonic antigen; Logistic regression; ROC curve

研究数据显示, 肺癌在人群死因中居各类恶性肿瘤之首^[1]。不论是作为肺癌的早期诊断指标, 还是作为治疗、判断预后的良好指标, 肿瘤标志物检测始终起着不可替代的作用。许多国内外学者对肺癌的肿瘤标志物做了大量有益的探索^[2-4]。根据 2003 年 WHO 对肺癌的分类, 将其分为鳞癌、腺癌、腺鳞癌、小细胞癌、大细胞癌、肉瘤样癌、类癌、唾液腺癌 8 个基本类型。由于肺癌有时与肺部良性疾病难以鉴别, 故易导致患者错过最佳治疗时期延误治疗。因此, 采用肿瘤标志物对肺癌的早期诊断具有很高的临床意义。癌胚抗原(CEA)、甲胎蛋白(AFP)、糖类抗原(CA)15-3、CA19-9、CA72-4 作为肿瘤标

志物已经在临床得到广泛应用^[5-7]。本研究通过收集石河子大学医学院第一附属医院肺癌患者、肺部良性疾病患者及健康人群血 CEA、AFP、CA15-3、CA19-9、CA72-4 的数据, 运用 Logistic 回归和受试者工作特征(ROC)曲线综合评价各血清指标在肺癌诊断中的价值。现报道如下。

1 资料与方法

1.1 一般资料 肺癌组病例为 2013 年 1 月到 2015 年 10 月在石河子大学第一附属医院住院治疗的患者, 共 62 例, 其中男 41 例, 女 21 例, 年龄 30~78 岁, 平均(61.0±10.4)岁。所有病例均经临床和病理确诊, 其中腺癌 37 例, 鳞癌 20 例和小细胞

* 基金项目: 国家自然科学基金资助项目(81160001); 石河子大学大学生创新训练计划项目(SRP2016060)。
作者简介: 陈佳, 男, 在读本科, 主要从事感染与免疫研究。 △ 通信作者, E-mail: 1556874645@qq.com。

癌 5 例。肺部良性疾病患者(均经临床或病理证实)共 58 例纳入肺部良性疾病组,其中男 37 例,女 21 例,年龄 16~79 岁,平均(59.0±11.1)岁。健康对照组为石河子大学第一附属医院体检的健康成人,共 59 例,其中男 45 例,女 14 例,年龄 19~75 岁,平均(60.0±9.9)岁。3 组的性别、年龄等一般资料比较,差异无统计学意义($P>0.05$),具有可比性。

1.2 方法 所有受检者清晨空腹抽取静脉血 3 mL,分离血清,采用 Roche E170 电化学发光仪测定血清 CEA、AFP、CA15-3、CA19-9、CA72-4 水平。所用检测试剂均为相应配套试剂,严格按照说明书操作。各指标的正常参考范围:CEA<3.4 μg/L,AFP<5.6 μg/L,CA15-3<25 μg/L,CA19-9<27 μg/L,CA72-4<6.9 μg/L。

1.3 统计学处理 采用 SPSS19.0 统计软件进行分析,Shapiro-Wilk 法进行正态性检验,Levene 法进行方差齐性检查,正态计量资料以 $\bar{x}\pm s$ 表示,各指标的组间比较采用 F 检验;非正态计量资料以 $M(P_{50}, P_{75})$ 表示,组间比较采用秩和检验。用

表 1 3 组各肿瘤标志物水平比较[$M(P_{50}, P_{75}), \mu\text{g/L}$]

组别	<i>n</i>	AFP	CEA	CA-153	CA19-9	CA72-4
健康对照组	59	1.85(1.34,2.50)*	1.81(1.29,2.80)*	8.68(6.36,11.92)*	9.46(7.38,14.70)*	1.35(0.93,2.49)*
肺部良性疾病组	58	1.72(1.04,2.51)*	1.36(0.92,1.84)*	10.01(7.81,13.40)*	6.46(3.19,11.74)*	1.34(1.06,2.17)*
肺癌组	62	1.78(1.27,2.82)	3.11(1.79,6.08)	14.15(9.11,23.50)	12.43(6.63,20.93)	2.15(1.55,6.47)

注:与肺癌组比较,* $P<0.05$ 。

表 2 4 种肿瘤标志物在不同病理类型肺癌中的水平[$M(P_{50}, P_{75}), \mu\text{g/L}$]

类别	<i>n</i>	CEA	CA-153	CA19-9	CA72-4
鳞癌	20	2.26(1.56,4.78)	10.73(8.75,26.14)	12.26(4.16,17.20)	2.03(1.43,3.17)
腺癌	37	3.23(2.51,6.24)	15.71(10.30,23.44)	12.63(8.36,26.72)	2.58(1.71,6.31)
小细胞癌	5	3.14(1.87,8.00)	22.34(10.76,34.80)	9.63(1.83,22.66)	2.68(1.49,39.41)

2.3 4 种肿瘤标志物对肺癌诊断的敏感性、特异性 以肺部良性疾病组为对照,根据公式,计算 CEA、CA15-3、CA19-9、CA72-4 对肺癌的敏感性分别是 45.16%(28/62)、17.74%(11/62)、17.74%(11/62)、24.19%(15/62),特异性分别是 98.28%(57/58)、98.28%(57/58)、96.55%(56/58)、93.10%(54/58),4 项联合检查的敏感性和特异性分别为 26.21%、96.48%。

2.4 4 种肿瘤标志物的回归分析 设 $X_1 = \text{CEA}, X_2 = \text{CA15-3}, X_3 = \text{CA19-9}, X_4 = \text{CA72-4}$,得肺癌预测概率值回归模型 $Y=1/[1+\text{EXP}(1.982X_1+3.383X_2+2.373X_3+2.167X_4-1.679)]$,生成一组新变量 Y 。经 Logistic 回归分析,4 种肿瘤标志物的 OR 均较高,表明 4 种标志物与肺癌的关联强度较大,且 CA15-3 的 OR 最高,说明该指标对肺癌的临床诊断价值较其余 3 个指标更大。见表 3。

表 3 肿瘤标志物的 Logistic 回归分析

标志物	β	<i>SE</i>	<i>Wald</i>	<i>P</i>	<i>OR</i>	95% <i>CI</i>
CEA	1.982	0.482	16.899	0.000	7.259	2.821~18.679
CA-153	3.383	1.104	9.388	0.002	29.462	3.384~256.52
CA19-9	2.373	0.866	7.510	0.006	10.725	1.965~58.527
CA72-4	2.167	0.651	11.067	0.001	8.728	2.435~31.281
常数	-1.679	0.245	47.015	0.000	0.187	—

注:—表示无数据。

Logistic 回归筛选变量并建立回归方程,对新变量及各单项指标进行 ROC 曲线分析。以 $P<0.05$ 为差异有统计学意义。

2 结果

2.1 3 组各肿瘤标志物水平比较 经秩和检验,健康对照组、肺部良性疾病组的 CEA、CA15-3、CA19-9、CA72-4、AFP 水平比较,差异无统计学意义($P>0.05$),从而排除肺部良性疾病对肿瘤标志物的干扰。再将两组分别与肺癌组进行比较,CEA、CA15-3、CA19-9、CA72-4 水平差异均有统计学意义($P<0.05$),提示肺癌与肿瘤标志物有关。因此,上述 4 种肿瘤标志物具有诊断价值。AFP 在 3 组的水平差异无统计学意义($P>0.05$),因此不作为诊断肺癌的肿瘤标志物。见表 1。

2.2 4 种肿瘤标志物在不同病理类型肺癌中的水平 经秩和检验,结果显示 CEA、CA15-3、CA19-9、CA72-4 在 3 种肺癌病理类型中的差异均无统计学意义($P>0.05$),因此,这 4 种肿瘤标志物用于临床肺癌的分型诊断,意义不大。见表 2。

表 4 4 种肿瘤标志物和新生产 Y 变量的 AUC

标志物	AUC	标准误	<i>P</i>	95% <i>CI</i>
CEA	0.783	0.038	0.000	0.708~0.857
CA-153	0.714	0.041	0.000	0.634~0.795
CA19-9	0.635	0.047	0.004	0.543~0.727
CA72-4	0.669	0.043	0.001	0.584~0.753
Y	0.794	0.040	0.000	0.716~0.872

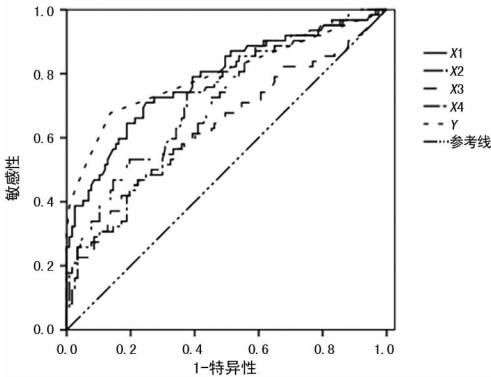


图 1 新变量 Y 的 ROC 曲线分析

2.5 肿瘤标志物和 Logistic 回归生成新变量 Y 的 ROC 曲线

分析 新生成的 Y 变量 (AFP+CA15-3+CA19-9+CA72-4) 的 ROC 曲线下面积 (AUC) 为 0.794, 明显大于任何单一肿瘤标志物的 AUC。见表 4 和图 1。

3 讨 论

肺癌的病死率很高, 居肿瘤疾病之首。因大部分肺癌患者被发现时已经处于晚期, 治疗存在一定的风险并且预后不佳。因此, 早发现、早诊断、早治疗是治疗肺癌的最佳选择。为了提高对肺癌诊断的准确性, 本研究使用 Logistic 回归和 ROC 曲线来分析血清中 5 种肿瘤标志物的表达水平与肺癌之间的相关性、敏感性、特异性, 为临床肺癌患者的早期诊断提供科学依据及理论参考依据。

AFP 是胚胎时期血液里含有的一种特殊蛋白质, 只有在胎儿的肝细胞中才能合成, 胎儿的这种蛋白质含量最高, 它在肝癌患者的血清中含量较高, 对肝癌诊断的阳性率达 90% 左右, 是肝癌最重要的肿瘤标志物。CEA 是一类具有人类胚胎抗原特异性决定簇的酸性糖蛋白, 属于非器官特异性肿瘤相关抗原, 分泌 CEA 的肿瘤大都位于空腔脏器。有研究报道, CEA 测定可以用于多种肿瘤的诊断, 其中包括肺癌的诊断^[8-9]。CA15-3 是一种由腺体分泌的多形上皮黏蛋白, 存在于多种腺癌内, 其中包括肺癌^[10]。CA19-9 是一种新的碳水化合物抗原, 即单涎酸神经节苷脂, 不但存在于胰管上皮中, 在肺癌疾病中也可呈现阳性^[11]。CA72-4 是一种不同于 CEA、CA19-9 和 CA15-3 的肿瘤相关蛋白, 属于黏蛋白类癌胚抗原, 于 1981 年从乳腺癌的肝转移灶中发现。蒋灵军等^[12]报道, CA72-4 在肺癌中具有较高的表达水平。

本研究显示, 肺癌患者血清 CEA、CA15-3、CA19-9、CA72-4 水平分别高于健康对照组和肺部良性疾病组 ($P < 0.05$), 提示这 4 种肿瘤标志物对肺癌有诊断价值。如果继续对肺癌肿瘤分期和分级进行研究, 可对临床疾病的进展和预后提供更多参考价值。但本研究未获得相关临床样本的分级与分期数据, 故未做相关研究。表 2 数据显示, CEA、CA15-3、CA19-9、CA72-4 在鳞癌、腺癌和小细胞性肺癌的差异无统计学意义 ($P > 0.05$), 表明以上指标分别对于鳞癌、腺癌及小细胞癌均不具有诊断及治疗指导意义。

本研究显示, CEA 的敏感度 (45.16%) 位居 4 种肿瘤标志物首位, 对于肺癌诊断的敏感性最高; 而 CA15-3 特异性 (98.28%) 最高, 但 CA15-3 的敏感度 (17.74%) 为 4 种肿瘤标志物中最低, 这与多数文献报道的, 提高了特异性却降低了敏感性的结论一致^[13-15]。Logistic 回归结果显示, CA15-3 与肺癌的关联强度最高, 而 CEA、CA15-3、CA19-9、CA72-4 组合的肿瘤标志物正确诊断指数最高。因此, 使用 Logistic 回归模型作为诊断模型, 具有较高的科学价值。不同变量得到的 ROC 曲线的 AUC 越大, 代表其临床准确性越高, 价值越大。参照 AUC 评价标准, AUC 在 0.5~<0.7 时有较低准确性, 在 0.7~0.9 时有一定准确性, 在 0.9 以上时有较高准确性。结合表 4 和图 1, 本研究中 4 种肿瘤标志物联合的新变量 Y 的 AUC 为 0.794, 相比各单一肿瘤标志物的 AUC 数值都大, 但只具有一定的准确性, 诊断意义不高。目前国际流行的肺癌肿瘤标志物检测如下, 细胞角蛋白 21-1 片段、胃泌素释放肽前体、抑癌基因 p53 的产物、神经元特异性烯醇化酶, 均具有较高的诊断准确性。建议医院将诊断指标试用为以上标志物, 以提

高对肺癌的诊断。

综上所述, 本研究运用 Logistic 回归和 ROC 曲线两种模型综合分析, 所得到的结论对于肺癌的诊断具有一定的实用价值, 对临床选择肺癌肿瘤标志物具有一定的指导意义。

参考文献

- [1] 姚晓军, 刘伦旭. 肺癌的流行病学及治疗现状[J]. 现代肿瘤医学, 2014, 21(8): 1982-1986.
- [2] 江涛. 比较 5 种血清肿瘤标志物联合检测对肺癌诊断和治疗的临床效果[J]. 中国医学创新, 2014, 6(3): 62-64.
- [3] 何全利, 吴艳艳, 鲁之中, 等. 五种血清肿瘤标志物联合检测对肺癌的临床诊断价值[J]. 中国临床研究, 2014, 27(2): 224-226.
- [4] 熊杰, 白生华, 徐万清, 等. 血清肿瘤标志物联合检测对肺癌的诊断价值[J]. 现代肿瘤医学, 2012, 20(3): 536-538.
- [5] 廖君群. 胸腔积液 CEA、CA19-9、CYFRA21-1、AFP 和 FERR 在肺癌诊断中的价值[J]. 国际检验医学杂志, 2013, 34(10): 1216-1218.
- [6] 宋清玲, 雷光文. SF、CEA、CA15-3、CA19-9、SCC 及 SF+CEA 联检对肺癌的诊断价值[J]. 放射免疫学杂志, 2011, 24(1): 118-119.
- [7] 朱海龙. CEA、AFP、CA125、CA15-3、CA19-9 检测对恶性肿瘤的诊断价值探讨[J]. 现代医药卫生, 2006, 22(22): 3422-3423.
- [8] 杜军华, 乔洪源, 尹宜发. 血清 CEA、CA125 及 Cyfra21-1 水平对中晚期非小细胞肺癌患者预后的影响[J]. 肿瘤防治研究, 2016, 43(2): 137-140.
- [9] 王媛, 张华丽. 肿瘤标志物 CEA、NSE、CA125、Pro GRP 在肺癌诊断中的价值[J]. 陕西医学杂志, 2016, 44(1): 14-15.
- [10] 聂书伟, 吕勇刚, 易军. 肺癌临床分期与肿瘤标志物 CEA、CA15-3、NSE 和 CYFRA21-1 相关性分析[J]. 临床军医杂志, 2012, 39(2): 395-398.
- [11] 张毅敏, 夏文进, 毛彩萍, 等. 血清标志物 CYFRA21-1、NSE、CEA、CA19-9、CA125、SCC 联合检测在肺癌诊断中的应用价值[J]. 中国癌症杂志, 2008, 18(4): 306-309.
- [12] 蒋灵军, 高树兴, 王玲珑, 等. 血清 CEA、CYFRA21-1、CA72-4、CA50 联检对肺癌诊断的探讨[J]. 放射免疫学杂志, 2011, 24(5): 525-527.
- [13] 马开慧, 金晔, 王伟. 四项肿瘤标志物联合检测在肺癌辅助诊断中的应用价值[J]. 中国实验诊断学, 2012, 16(12): 2288-2289.
- [14] 王磊, 何洁, 姚勇, 等. 血清肿瘤标志物联合检测对肺癌的诊断价值[J]. 标记免疫分析与临床, 2013, 20(3): 140-142.
- [15] 陶俊, 江莲, 唐良法. 肿瘤标志物 SCC-Ag、NSE、CEA 联合检测对肺癌诊断的价值[J]. 南通大学学报(医学版), 2010, 30(6): 436-438.