

• 论著 •

医疗大数据对老年肺炎患者预后的预测价值 ——基于北京市朝阳区医院医联体朝阳急诊病房数据结果

李鹏¹ 张兴厅² 尹芳² 郭俐彤² 马超² 才红波² 郭树彬¹

¹首都医科大学附属北京朝阳医院急诊科, 北京 100020; ²联仁健康医疗大数据科技股份有限公司, 北京 100013

通信作者: 郭树彬, Email: shubinguo@126.com

【摘要】 目的 探索一种利用医疗大数据算法筛选临床数据库中能够用于评估老年肺炎患者预后的核心指标。**方法** 基于首都医科大学附属北京朝阳医院医联体朝阳急诊病房临床数据库, 应用大数据检索技术, 以数据库中老年肺炎患者为研究对象, 根据出院时预后将患者分为死亡组和存活组。收集患者的一般资料, 包括性别、年龄、血气、实验室指标集合数据, 使用计算机语言 Python 批量计算出影响老年肺炎患者死亡的关键指标, 并采用 Logistic 回归分析实验室指标与患者预后的相关性; 绘制受试者工作特征曲线 (ROC 曲线), 分析本研究使用的筛选方法对患者预后的预测价值。**结果** 最终入选 265 例患者, 死亡 64 例, 存活 201 例。取每例患者入院首次检测指标的数据, 最终从 472 项指标中筛选出 23 项差异有统计学意义的关键指标, 包括: 血常规指标 7 项、血气指标 3 项、肿瘤标志物指标 3 项、凝血功能相关指标 4 项、营养及器官功能相关指标 6 项。① 肺炎死亡患者血气关键指标: 51.6% (33 例) 的患者 Cl^- 浓度为 $97 \sim 111 \text{ mmol/L}$, 81.2% (52 例) 的患者血乳酸 (Lac) 为 $0.5 \sim 2.5 \text{ mmol/L}$, 87.5% (56 例) 的患者 H^+ 浓度为 $0 \sim 46 \text{ mmol/L}$ 。② 肺炎死亡患者血常规关键指标: 46.9% (30 例) 的患者血红蛋白 (Hb) 为 $80 \sim 109 \text{ g/L}$, 67.2% (43 例) 的患者血中嗜酸粒细胞比例 (EOS%) 为 $0.000 \sim 0.009$, 51.6% (33 例) 的患者血中淋巴细胞比例 (LYM%) 为 $0.00 \sim 0.09$, 50.0% (32 例) 的患者血中红细胞计数 (RBC) 为 $(3.0 \sim 3.9) \times 10^{12}/L$, 54.7% (35 例) 的患者血中白细胞计数 (WBC) 为 $(0.0 \sim 9.9) \times 10^9/L$, 48.4% (31 例) 的患者血中红细胞分布宽度变异系数 (RDW-CV) 为 $10.0\% \sim 14.9\%$, 48.4% (31 例) 的患者血中 C-反应蛋白 (CRP) 为 $0.0 \sim 49.9 \text{ mg/L}$ 。③ 肺炎死亡患者肿瘤标志物关键指标: 76.6% (49 例) 的患者血清游离前列腺特异抗原 / 总前列腺特异抗原 (FPSA/TPSA) 为阴性 (比值为 0), 92.2% (59 例) 的患者细胞角蛋白 19 片段 (CYFRA21-1) 为 $0.0 \sim 11.0 \mu\text{g/L}$, 75.0% (48 例) 的患者糖类抗原 125 (CA125) 为 $0 \sim 104 \text{ kU/L}$ 。④ 肺炎死亡患者凝血功能关键指标: 68.8% (44 例) 的患者活化部分凝血活酶时间 (APTT) 为 $57 \sim 96 \text{ s}$, 73.4% (47 例) 的患者 D-二聚体为 $0 \sim 6 \text{ mg/L}$, 93.8% (60 例) 的患者凝血酶时间 (TT) 为 $14 \sim 22 \text{ s}$, 89.1% (57 例) 的患者二磷酸腺苷 (ADP) 的抑制率为 $0\% \sim 53\%$ 。⑤ 肺炎死亡患者营养及器官功能关键指标: 92.2% (59 例) 的患者 B 型脑钠肽 (BNP) 为 0, 46.9% (30 例) 的患者前白蛋白 (PA) 为 $71 \sim 140 \text{ mg/L}$, 90.6% (58 例) 的患者尿酸 (UA) 为 $21 \sim 41 \mu\text{mol/L}$, 75.0% (48 例) 的患者白蛋白 (Alb) 为 $10 \sim 20 \text{ g/L}$, 93.5% (60 例) 患者白蛋白 / 球蛋白比值 (A/G 比值) 为 $0 \sim 0.9$, 84.4% (54 例) 的患者乳酸脱氢酶 (LDH) 为 $0 \sim 6.68 \mu\text{mol/L} \cdot \text{s}^{-1} \cdot \text{L}^{-1}$ 。⑥ Logistic 回归和 ROC 曲线分析: Logistic 回归分析表明, PA 和 Lac 是影响患者预后的因素, PA 可使死亡风险降低 0.9%, Lac 可使死亡风险增加 69.4%; 实验室指标与患者死亡预测模型预测效果的 ROC 曲线下面积 (AUC) = 0.80, 说明本研究使用的筛选方法效果较好, 通过本研究模型能较好地预测老年肺炎患者预后。**结论** 运用大数据技术可从急诊病房临床数据库中筛选出 23 项用于评估老年肺炎患者预后的核心指标, 为临床评估老年肺炎患者预后提供了新的角度和方法。

【关键词】 大数据; 数据挖掘; 急诊; 老年肺炎

DOI: 10.3760/cma.j.cn121430-20200611-00461

The predictive value of medical big data for the prognosis of elderly patients with pneumonia: based on the result of clinical database of a Beijing Chaoyang Hospital Consortium Chaoyang Emergency Ward

Li Peng¹, Zhang Xingting², Yin Fang², Guo Litong², Ma Chao², Cai Hongbo², Guo Shubin¹

¹Department of Emergency, Beijing Chaoyang Hospital, Capital Medical University, Beijing 100020, China; ²Lianren Healthcare Big Data Technology Co., Ltd, Beijing 100013, China

Corresponding author: Guo Shubin, Email: shubinguo@126.com

【Abstract】 Objective To explore a medical big data algorithm to screen the core indicators in clinical database that can be used to evaluate the prognosis of elderly patients with pneumonia. **Methods** Based on the clinical database of a Beijing Chaoyang Hospital Consortium Chaoyang Emergency Ward in Beijing Chaoyang Hospital, Capital Medical University, patients with pulmonary infection were selected through the big data retrieval technology. According to the prognosis at the time of discharge, they were divided into death group and survival group. The general data of patients were collected, including gender, age, blood gas and laboratory indices. A computer language called Python was used to make batch calculations of key indicators that affect mortality in elderly patients with pneumonia. Logistic regression analysis was used to analyze the relationship between laboratory indicators and patients' prognosis. Receiver

operator characteristic curve (ROC curve) was drawn to analyze the predictive value of screening method for patients' prognosis. **Results** A total of 265 patients were included in the study, 64 died and 201 survived. The data of the first detection indexes of each patient after admission were collected, and 23 key indicators with significant differences were selected from 472 indicators: blood routine indicators ($n = 7$), blood gas indicators ($n = 3$), tumor markers indicators ($n = 3$), coagulation related indicators ($n = 4$), and nutrition and organ function indicators ($n = 6$). ① The key indicators of blood gas in patients died of pneumonia: Cl^- was 97–111 mmol/L in 51.6% (33 cases) of patients, lactic acid (Lac) was 0.5–2.5 mmol/L in 81.2% (52 cases) of patients, and H^+ was 0–46 mmol/L in 87.5% (56 cases) of patients. ② The key indicators of blood routine of patients died of pneumonia: hemoglobin count (Hb) of 46.9% (30 cases) patients was 80–109 g/L, the eosinophils proportions (EOS%) in 67.2% (43 cases) patients was 0.000–0.009, the lymphocytes proportions (LYM%) in 51.6% (33 cases) patients was 0.00–0.09, the red blood cell count (RBC) in 50.0% (32 cases) patients was $(3.0\text{--}3.9) \times 10^{12}/\text{L}$, the white blood cell count (WBC) in 54.7% (35 cases) patients was $(0.0\text{--}9.9) \times 10^9/\text{L}$, and the red blood cell volume distribution width coefficient of variability (RDW-CV) in 48.4% (31 cases) patients was 10.0%–14.9%, serum C-reactive protein (CRP) was 0.0–49.9 mg/L in 48.4% (31 cases) patients. ③ The key indicators of tumor markers in patients died of pneumonia: 76.6% (49 cases) of patients had negative free prostate specific antigen/total prostate specific antigen (FPSA/TPSA, the ratio was 0), 92.2% (59 cases) had cytokeratin 19 fragment (CYFRA21-1) between 0.0–11.0 $\mu\text{g}/\text{L}$, and 75.0% (48 cases) had carbohydrate antigen 125 (CA125) between 0–104 kU/L. ④ The key coagulation indexes of patients died of pneumonia: 68.8% (44 cases) of patients had activated partial thromboplastin time (APTT) of 57–96 s, 73.4% (47 cases) of patients had D-dimer of 0–6 mg/L, 93.8% (60 cases) of patients had thrombin time (TT) of 14–22 s, and 89.1% (57 cases) of patients had adenosine diphosphate (ADP) inhibition rate of 0%–53%. ⑤ Nutrition and organ function key indicators in patients died of pneumonia: 92.2% (59 cases) of brain natriuretic peptide (BNP) in patients with 0, 46.9% (30 cases) of patients had prealbumin (PA) of 71–140 mg/L, 90.6% (58 cases) of the patients with uric acid (UA) for 21–41 $\mu\text{mol}/\text{L}$, 75.0% (48 cases) of the patients with albumin (Alb) to 10–20 g/L, 93.5% (60 cases) of patients had albumin/globulin ratio (A/G ratio) of 0–0.9, 84.4% (54 cases) of the patients with lactate dehydrogenase (LDH) from 0–6.68 $\mu\text{mol}/\text{L} \cdot \text{s}^{-1} \cdot \text{L}^{-1}$. ⑥ Logistic regression analysis and ROC curve analysis: Logistic regression analysis showed that PA and Lac were the prognostic factors. PA could reduce the risk of death by 0.9%, Lac could increase the risk of death by 69.4%; the area under ROC curve (AUC) between laboratory indicators and the prediction effect of death prediction model for patients' prognosis was 0.80, which showed that the classification effect was better, and this study model could better predict the prognosis of elderly patients with pneumonia. **Conclusion** By using big data technology, 23 core indicators for evaluating the prognosis of elderly patients with pneumonia can be screened from the clinical database of emergency ward, which provides a new perspective and method for clinical evaluation of the prognosis of elderly patients with pneumonia.

【Key words】 Big data; Data mining; Emergency department; Senile pneumonia

DOI: 10.3760/cma.j.cn121430-20200611-00461

近年来,随着医院信息化基础设施的日益完善,医院信息系统(hospital information system, HIS)、电子病历(electronic medical record, EMR)系统、影像采集与传输系统(picture archiving and communication system, PACS)、实验室信息管理系统(laboratory information management system, LIMS)、病理信息系统(pathology information system, PIS)已经成为中国大多数医院的标准配备。一方面,海量病历、影像学资料、检验检查报告、财务管理等以电子化数据的形式被记录下来,经过 10 年甚至更长时间的积累,国内大型医院的年数据积累量已经达到 100 T 级别。另一方面,云计算、物联网、人工智能、5 G 等新一代技术在健康医疗领域的应用不断深入,使得数据来源和采集方式变得多种多样,健康医疗数据在类型上越来越复杂多样,在数据规模上更是爆发式增长,健康医疗大数据已逐步形成。

急诊科是医院内医疗设备最多的科室,也是各种数据的交汇地,集合了大量患者的体征数据、图像连续监测数据、设备监测数据、病历数据(电子病

历、检验信息等),这些数据维度多、时效性强、价值密度大、质量高,为临床科研提供了许多的原始素材。而如何运用大数据的思维开展急诊领域临床研究,提升急诊临床科研效率成为我们亟需研究的命题。本院急诊科经过多年发展,已积累了大量急诊病历数据,初步建立了急危重症大数据库,本研究基于数据库中 2017 年 1 月 1 日至 2018 年 12 月 31 日收治的 1 615 例入住首都医科大学附属北京朝阳医院医联体朝阳急诊病房、入院第一诊断明确为肺炎的老年患者为研究对象。收集患者在住院期间的所有实验室指标,应用大数据技术挖掘能用于评估老年肺炎死亡风险的核心指标,以期为进一步分析影响急诊科老年肺炎患者的死亡因素提供线索。

1 资料及方法

1.1 研究对象及方法: 2017 年 1 月 1 日至 2018 年 12 月 31 日首都医科大学附属北京朝阳医院医联体朝阳急诊病房共收录 1 615 例患者的临床资料,收集的实验室数据共有 14 个字段、1 085 882 条记录。通过对诊断结果进行统计分析显示,最多的为肺炎

患者,共有 435 例;其中男性 258 例,女性 177 例,考虑到研究对象年龄分布不均匀且偏大,根据数据特点及以往研究惯例^[1-2],选择 60 岁以上人群作为研究对象进行分析,最终纳入 265 例患者。

1.2 伦理学:本研究符合医学伦理学标准,数据收集取得医院科研部门审核批准(审批号:2020-03-10)。

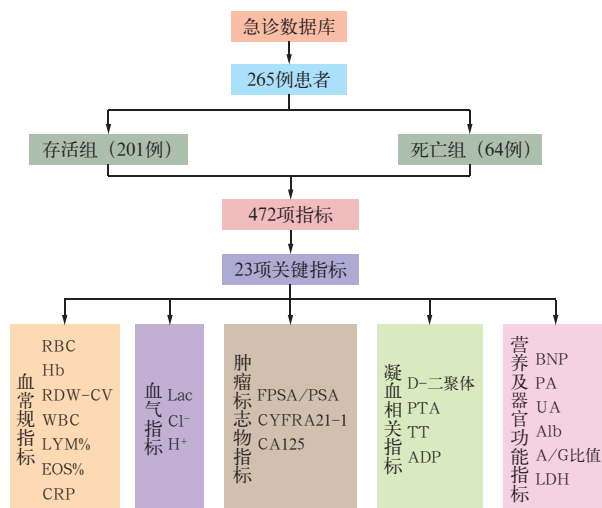
1.3 指标收集:收集患者一般资料,包括性别、年龄、血气、实验室指标集合数据,将这些数据分为血常规关键指标、凝血功能关键指标和生化关键指标,为了便于对照研究,对于连续测量指标只取首次和末次数据,入院第 1 次的指标定义为首次指标,出院前最后 1 次的指标定义为末次指标,对于住院期间仅测量 1 次的指标,首次末次按同值处理。

1.4 数据处理:数据处理是业务数据成为研究数据的最重要环节,主要有以下几个步骤:第一步,数据校对,对明显的错误数据通过溯源追踪的方法进行校正;第二步,数据合并,对于多次反复住院的患者进行合理合并,将住院时间处理后共收集 265 例患者;第三步,缺失值处理,由于原始数据中无样本来源的字段,通过业务系统和专业判断补全该字段,删除无法补全的条目;第四步,确定的异常值含义并把异常值的分类变量转化为可计算的数值变量。经过一系列数据处理步骤后,原始的 108 万余行数据减少为约 36 万行,真正利用到分析中的数据为 14 万行。

1.5 统计学处理:通过 HIS 和 LIMS 系统抽取相关数据导出到 Excel 中,运用计算机语言 Python 软件调用统计函数包进行批量数据处理后经 SPSS 18.0 统计软件验证。计数资料以例(%)表示。采用 Logistic 回归分析实验室指标与患者预后的相关性;绘制受试者工作特征曲线(receiver operator characteristic curve, ROC 曲线)评价死亡预测模型对老年肺炎患者预后的预测效果。 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 一般资料(图 1):取每例患者入院首次指标数据,对最终入选的 265 例患者按照结局分为死亡组(64 例)和存活组(201 例)。对每项指标分别进行差异性 t 检验,最终从 472 项指标中筛选出 23 项差异有统计学意义的关键指标,并将此 23 项关键指标归纳为 5 类:血常规指标 7 项、血气指标 3 项、肿瘤标志物指标 3 项、凝血功能相关指标 4 项、营养及器官功能指标 6 项。



注:RBC 为红细胞计数,Hb 为血红蛋白,RDW-CV 为红细胞分布宽度变异系数,WBC 为白细胞计数,LYM% 为淋巴细胞比例,EOS% 为嗜酸粒细胞比例,CRP 为 C-反应蛋白,Lac 为血乳酸,F-PSA/PSA 为游离前列腺特异抗原/总前列腺特异抗原,CYFRA21-1 为细胞角蛋白 19 片段,CA125 为糖类抗原 125,PTA 为凝血酶原活动度,TT 为凝血酶时间,ADP 为二磷酸腺苷,BNP 为 B 型脑钠肽,PA 为前白蛋白,UA 为尿酸,Alb 为白蛋白,A/G 比值为白蛋白/球蛋白比值,LDH 为乳酸脱氢酶

图 1 首都医科大学附属北京朝阳医院 2017 年 1 月 1 日至 2018 年 12 月 31 日收治的老年肺炎患者关键指标筛选流程

2.2 实验室指标分析

2.2.1 老年肺炎死亡患者血气关键指标(表 1):51.6%(33 例)的患者 Cl^- 为 97 ~ 111 mmol/L,81.2%(52 例)的患者血乳酸(lactic acid, Lac)为 0.5 ~ 2.5 mmol/L,87.5%(56 例)的患者 H^+ 为 0 ~ 46 mmol/L。

表 1 首都医科大学附属北京朝阳医院 2017 年 1 月 1 日至 2018 年 12 月 31 日收治 64 例老年肺炎死亡患者血气关键指标分布情况

指标	值域	例数 [例(%)]	指标	值域	例数 [例(%)]
Cl^- (mmol/L)	0 ~ 81	3 (4.7)	H^+ (mmol/L)	2.6 ~ 3.5	8 (12.5)
	82 ~ 96	10 (15.6)		≥ 3.6	4 (62.5)
	97 ~ 111	33 (51.6)	Lac (mmol/L)	0 ~ 26	27 (42.1)
	112 ~ 126	15 (23.4)		27 ~ 46	29 (45.3)
	127 ~ 141	3 (4.7)		47 ~ 66	4 (62.5)
Lac (mmol/L)	0.5 ~ 1.5	32 (50.0)		67 ~ 86	3 (4.7)
	1.6 ~ 2.5	20 (31.2)		≥ 86	1 (15.6)

注:Lac 为血乳酸

2.2.2 老年肺炎死亡患者血常规关键指标分析(表 2):46.9%(30 例)的患者血红蛋白(hemoglobin, Hb)为 80 ~ 109 g/L,67.2%(43 例)的患者嗜酸粒细胞比例(eosinophils proportions, EOS%)为 0.000 ~ 0.009,51.6%(33 例)的患者淋巴细胞比例(lymphocyte proportions, LYM%)为 0.00 ~ 0.09,50.0%(32 例)的患者红细胞计数(red blood cell count, RBC)为 $(3.0 \sim 3.9) \times 10^{12}/L$,54.7%(35 例)的患者白细胞计数(white

blood cell count, WBC) 为 $(0.0 \sim 9.9) \times 10^9/L$, 48.4% (31 例) 的患者红细胞分布宽度变异系数 (red blood cell volume distribution width coefficient of variability, RDW-CV) 为 10.0% ~ 14.9%, 48.4% (31 例) 的患者 C-反应蛋白 (C-reactive protein, CRP) 为 0.0 ~ 49.9 mg/L。

表 2 首都医科大学附属北京朝阳医院 2017 年 1 月 1 日至 2018 年 12 月 31 日收治 64 例肺炎死亡患者血常规关键指标分布情况

指标	值域	例数 [例(%)]	指标	值域	例数 [例(%)]
Hb(g/L)	0 ~ 49	1 (1.6)		2.0 ~ 2.9	21 (32.8)
	50 ~ 79	16 (25.0)		3.0 ~ 3.9	32 (50.0)
	80 ~ 109	30 (46.9)		4.0 ~ 4.9	8 (12.5)
	110 ~ 139	15 (23.4)	WBC	0.0 ~ 9.9	35 (54.7)
	140 ~ 169	2 (3.1)		($\times 10^9/L$) 10.0 ~ 19.9	22 (34.4)
EOS%	0.000 ~ 0.009	43 (67.2)		20.0 ~ 29.9	3 (4.7)
	0.010 ~ 0.019	10 (15.6)		30.0 ~ 39.9	2 (3.1)
	0.020 ~ 0.029	6 (9.4)	RDW-CV	0.0 ~ 9.9	1 (1.6)
	0.030 ~ 0.039	5 (7.8)		(%) 10.0 ~ 14.9	31 (48.4)
LYM%	0.00 ~ 0.09	33 (51.6)		15.0 ~ 19.9	29 (45.3)
	0.10 ~ 0.19	19 (29.7)		20.0 ~ 24.9	3 (4.7)
	0.20 ~ 0.29	7 (10.9)	CRP(mg/L)	0.0 ~ 49.9	31 (48.4)
	≥ 0.30	5 (7.8)		50.0 ~ 99.9	20 (1.2)
RBC	0.0 ~ 0.9	1 (1.6)		100.0 ~ 149.9	9 (14.1)
	($\times 10^{12}/L$) 1.0 ~ 1.9	2 (3.1)		150.0 ~ 199.9	4 (6.2)

注: Hb 为血红蛋白, EOS% 为嗜酸粒细胞比例, LYM% 为淋巴细胞比例, RBC 为红细胞计数, WBC 为白细胞计数, RDW-CV 为红细胞分布宽度变异系数, CRP 为 C-反应蛋白

2.2.3 老年肺炎死亡患者肿瘤标志物关键指标 (表 3): 76.6% (49 例) 的患者血清游离前列腺 / 总前列腺特异抗原 (free prostate specific antigen/total prostate specific antigen, FPSA/TPSA) 为阴性 (比值为 0), 92.2% (59 例) 的患者细胞角蛋白 19 片段 (cytokeratin 19 fragment antigen 21-1, CYFRA21-1) 为 0.0 ~ 11.0 $\mu g/L$, 75.0% (48 例) 的患者糖类抗原 125 (carbohydrate antigen, CA125) 为 0 ~ 104 kU/L。

表 3 首都医科大学附属北京朝阳医院 2017 年 1 月 1 日至 2018 年 12 月 31 日收治 64 例肺炎死亡患者肿瘤标志物关键指标分布情况

指标	值域	例数 [例(%)]	指标	值域	例数 [例(%)]
FPSA/TPSA	0.0 ~ 0.1	49 (76.6)		22.0 ~ 31.0	1 (1.6)
	0.2 ~ 10.1	4 (6.2)		≥ 31	1 (1.6)
	10.2 ~ 20.1	6 (9.4)	CA125	0 ~ 4	22 (34.4)
	20.2 ~ 30.1	3 (4.7)		(kU/L) 5 ~ 104	26 (40.6)
	≥ 30.1	2 (3.1)		105 ~ 204	9 (14.1)
CYFRA	0.0 ~ 0.9	21 (32.8)		205 ~ 304	5 (7.8)
	($\mu g/L$) 1.0 ~ 11.0	38 (59.4)		305 ~ 404	1 (1.6)
	12.0 ~ 21.0	3 (4.7)		405 ~ 504	1 (1.6)

注: FPSA/TPSA 为游离前列腺特异抗原 / 总前列腺特异抗原, CYFRA21-1 为细胞角蛋白 19 片段, CA125 为糖类抗原 125

2.2.4 老年肺炎死亡患者凝血功能关键指标 (表 4): 68.8% (44 例) 的患者活化部分凝血活酶时间 (activated partial thromboplastin time, APTT) 为 57 ~ 96 s, 73.4% (47 例) 的患者 D-二聚体为 0 ~ 6 mg/L, 93.8% (60 例) 的患者凝血酶时间 (thrombin time, TT) 为 14 ~ 22 s, 89.1% (57 例) 患者二磷酸腺苷 (adenosine diphosphate, ADP) 抑制率为 0% ~ 53%。

表 4 首都医科大学附属北京朝阳医院 2017 年 1 月 1 日至 2018 年 12 月 31 日收治 64 例肺炎死亡患者凝血功能关键指标分布情况

指标	值域	例数 [例(%)]	指标	值域	例数 [例(%)]
APTT(s)	0 ~ 36	1 (1.6)	TT(s)	< 14	1 (1.6)
	37 ~ 56	7 (10.9)		14 ~ 18	38 (59.4)
	57 ~ 76	22 (34.4)		19 ~ 22	22 (34.4)
	77 ~ 96	22 (34.4)		23 ~ 26	1 (1.6)
	97 ~ 116	9 (14.1)		≥ 26	2 (3.1)
	≥ 116	3 (4.7)	ADP 抑制率	0 ~ 53	57 (89.1)
D-二聚体	0 ~ 3	32 (50.0)		(%) 54 ~ 63	2 (3.1)
	(mg/L) 4 ~ 6	15 (23.4)		64 ~ 73	1 (1.6)
	7 ~ 9	8 (12.5)		74 ~ 93	3 (4.7)
	10 ~ 12	2 (3.1)		94 ~ 103	1 (1.6)
	≥ 12	7 (10.9)			

注: APTT 为活化部分凝血活酶时间, TT 为凝血酶时间, ADP 为二磷酸腺苷

2.2.5 老年肺炎死亡患者营养及器官功能关键指标 (表 5): 92.2% (59 例) 的患者 B 型脑钠肽 (brain natriuretic peptide, BNP) 为 0, 46.9% (30 例) 的患者前白蛋白 (prealbumin, PA) 为 71 ~ 140 mg/L, 90.6% (58 例) 的患者尿酸 (uric acid, UA) 为 21 ~ 41 $\mu mol/L$, 75.0% (48 例) 的患者白蛋白 (albumin, Alb) 为 10 ~ 20 g/L, 93.8% (60 例) 的患者白蛋白 / 球蛋白比值 (albumin globulin ratio, A/G ratio) 为 0 ~ 0.9, 87.5% (56 例) 的患者乳酸脱氢酶 (lactate dehydrogenase, LDH) 为 0 ~ 6.68 $\mu mol \cdot s^{-1} \cdot L^{-1}$ 。

2.3 影响老年肺炎患者预后的实验室指标与预后的 Logistic 回归分析 (表 6): 为进一步分析实验室指标与患者预后的关系, 对与本院收治老年肺炎患者相关性最大的前 10 项实验室指标 (相关系数绝对值 0.4 以上) 进行 Logistic 回归分析, 进而可以列出逻辑回归方程 ($Y=0.004 \times Hb+0.162 \times HCT+0.009 \times LYM\%+0.694 \times Lac+0.01 \times Alb-0.267 \times EOS\%-0.009 \times PA-2.186 \times RBC-0.064 \times WBC+0.164$), 用来预测患者死亡风险, 结果显示, PA 和 Lac 是影响患者预后的因素, PA 可使死亡风险降低 0.9%, Lac 可使死亡风险增加 69.4%。

2.4 ROC 曲线分析(图 2): 为评价死亡预测模型的预测效果,绘制 ROC 曲线并计算曲线下面积(area under ROC curve, AUC)=0.80,说明分类效果较好,通过以上模型能较好地预测老年肺炎患者预后。

表 5 首都医科大学附属北京朝阳医院 2017 年 1 月 1 日至 2018 年 12 月 31 日收治 64 例老年肺炎死亡患者营养及器官功能关键指标分布情况

指标	值域	例数 〔例(%)〕	指标	值域	例数 〔例(%)〕
BNP	0	59(92.2)	Alb(g/L)	0~9	12(18.8)
(mg/L)	1~50	2(3.1)		10~20	48(75.0)
	>51	3(4.7)		21~30	3(4.7)
PA(mg/L)	0	9(14.1)		31~40	1(1.6)
	1~70	9(14.1)	A/G 比值	<0.1	24(37.5)
	71~140	30(46.9)		0.1~0.5	20(31.2)
	141~210	11(17.2)		0.6~0.9	16(25.0)
	211~280	4(6.2)		1.0~1.3	4(6.2)
	281~350	1(1.6)	LDH	<0.02	18(28.1)
UA	<21	1(1.6)	($\mu\text{mol} \cdot \text{s}^{-1} \cdot \text{L}^{-1}$)	0.02~3.35	15(23.4)
($\mu\text{mol/L}$)	21~31	34(53.1)		3.36~6.68	23(35.9)
	32~41	24(37.5)		6.69~10.02	4(6.2)
	42~51	5(7.8)		10.03~13.36	2(3.1)
				13.37~16.70	1(1.6)
				≥ 16.70	1(1.6)

注:BNP 为 B 型脑钠肽,PA 为前白蛋白,UA 为尿酸,Alb 为白蛋白,A/G 比值为白蛋白/球蛋白比值,LDH 为乳酸脱氢酶

表 6 影响老年肺炎患者预后的实验室指标与预后 Logistic 回归分析结果

变量	β 值	t 值	P 值	变量	β 值	t 值	P 值
Hb	0.004	0.098	0.922	Alb	0.010	0.323	0.747
HCT	0.162	1.170	0.243	RBC	-2.186	-2.783	0.006
EOS%	-0.267	-1.824	0.069	WBC	0.064	1.989	0.047
PA	-0.009	-3.192	0.002	Mb	-0.000	-0.201	0.841
LYM%	0.009	0.505	0.614	截距	0.164	0.147	0.883
Lac	0.694	3.179	0.002				

注:Hb 为血红蛋白,HCT 为血细胞比容,EOS% 为嗜酸粒细胞比例,PA 为前白蛋白,LYM% 为淋巴细胞比例,Lac 为血乳酸,Alb 为白蛋白,RBC 为红细胞计数,WBC 为白细胞计数,Mb 为肌红蛋白

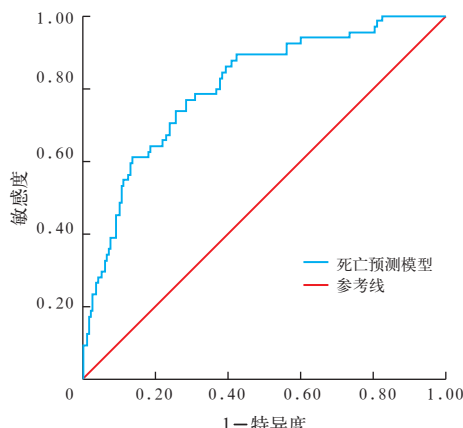


图 2 死亡预测模型对老年肺炎患者预后预测值的受试者工作特征曲线(ROC 曲线)

3 讨论

深入挖掘医疗大数据的潜在价值为临床医学服务,目前已逐渐成为医学界的共识,国内外学者已开展了许多基于大数据在急诊临床应用方面的研究。Janke 等^[3]探索了急诊领域大数据的预测分析潜力,通过挖掘海量电子病历数据可使急诊诊断更加精准,急诊资源利用更加高效。Ram 等^[4]利用环境、社交及网络等大数据预测急诊哮喘患者访问量准确率达到了 70%。Taylor 等^[5]利用本地大数据驱动的机器学习方法预测急诊科脓毒症患者的院内病死率,该方法优于传统基于规则的方法和医院现有预测技术。朱志勇等^[6]分析了急诊大数据的特点并提出重症监护病房(intensive care unit,ICU)转移、ICU 死亡和感染性休克 3 个预警模型,展示了这些模型在临床实验上的结果。贾立静等^[7]应用大数据算法筛选急救数据库中能用于评估急诊致命性消化道院内再出血的核心指标,为临床诊断该病提供了新的思路和方法。赵宇卓等^[8]分析当前急救大数据平台建设与临床决策支持系统的建设现状,以达到推动国内急救大数据发展的目的。

已有较多研究使用传统模式探索过影响老年肺炎患者的死亡因素,然而现有研究通常存在观察指标数目较为局限、样本量较少等问题^[9-11],而基于大数据分析技术,应用创新算法全面系统提取与研究结局相关核心指标的方法能有效解决以上问题。本研究利用大数据的处理方法,通过多种融合算法,对千万条数据批量分析,对急诊肺炎患者的 472 项检验指标进行了筛选,发现 23 项指标会影响到急危重症患者的临床预后,其中血常规指标 7 项,血气指标 3 项,肿瘤标志物指标 3 项,凝血功能相关指标 4 项,营养及器官功能指标 6 项,并将上述指标分成 3 类:第一类为目前已知与感染及死亡有明确直接或间接关系的指标,共 8 项,包括:WBC、LYM%、EOS、CRP、Lac、H⁺、PA、Alb;第二类为可能与死亡有关联的指标,共 11 项,包括:血 RBC、Hb、FPSA/TPSA、CYFRA21-1、CA125、ADP 抑制率、APTT、D-二聚体、TT、BNP、A/G 比值;第三类为初次发现或以前被忽视或未被重视的与死亡相关的指标,共 4 项,包括:血 RDW-CV、Cl⁻、UA、LDH。对于第一类为医学界有明确共识的相关指标,老年肺炎属于感染性疾病,与感染相关的指标一定与预后相关,值得注意的是不少学者强调患者内环境情况和营养状况与预后也同样有着不可忽视的关系,通

过大数据的“真实世界”证据更加强有力地支持了这种观点。第二和第三类指标才是真正体现大数据优势的地方,第二类指标定义为有争议或根据经验有倾向性结论的但未经大规模论证的指标,第三类为既往忽略的或未被认识到的指标。通过对庞大数据的整理、分析从中挖掘出差异有统计学意义的指标,对照以往的研究结论经过筛选分类出第二类和第三类指标,本研究第二类指标有 11 项,第三类指标有 4 项。本试验可为第二类指标相关的研究提供证据支持,亦提示这类指标确实存在实际临床意义。第三类指标则给人们提供了某些新的线索或视角,也许为今后研究某些问题开拓了一条新的思路。本研究关于肺炎的数据分析第二类指标有 11 项,人们可据此查找文献,整理以往数据,重新认识这类指标的真实意义。本研究选取相关性最大的前 10 项指标(相关系数绝对值 0.4 以上)进行 Logistic 回归分析,并计算出 $AUC=0.80$,说明死亡预测模型能较好地地区分死亡和存活患者,较为准确地预测老年肺炎患者的预后,利用该模型对急诊老年肺炎患者进行预后风险评估,提醒医生尽早对高风险患者采取更积极的治疗措施,对于降低老年肺炎住院患者病死率有重要意义。

综上所述,既往对于肺炎患者终点事件的预测指标有很多,但由于数据量和样本量的限制,存在混杂和偏倚的可能,通过大数据的应用就能相对全面地找出关键性指标进行量化分析,给出最佳模型,指导临床实践,这些都是传统医学研究的局限。随着计算设备的升级,算法的改进,算力的提升,就可以随时更新数据,分析数据,以指导临床实践,真正发挥大数据在医疗领域的重要作用。本研究初步证明基于真实临床业务产生的医疗大数据在临床研究中的价值,是以“真实世界”数据解决临床实际问题的有益探索,为未来急危重症患者的临床研究提供了新的角度和方法。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 饶芝国,刘宗明,林贵军,等.老年脑卒中患者医院内肺部感

染的危险因素分析[J].中华老年医学杂志,2008,27(7):518-520.

Rao ZG, Liu ZM, Lin GJ, et al. Risk factors for nosocomial pneumonia in elderly stroke patients[J]. Chin J Geriatr, 2008, 27(7): 518-520.

[2] 刘新海,曾凡华,叶俊英,等.不同年龄段肺炎患者肺炎病原菌分布及耐药性情况分析[J].河北医学,2014,20(7):1074-1076. DOI: 10.3969/j.issn.1006-6233.2014.07.006.

Liu XH, Zeng FH, Ye JY, et al. Analysis on the distribution of pathogenic bacteria and drug resistance situation of patients with pneumonia at different ages[J]. Hebei Med, 2014, 20(7): 1074-1076. DOI: 10.3969/j.issn.1006-6233.2014.07.006.

[3] Janke AT, Overbeek DL, Koehler KE, et al. Exploring the potential of predictive analytics and big data in emergency care[J]. Ann Emerg Med, 2016, 67(2): 227-236. DOI: 10.1016/j.annemergmed.2015.06.024.

[4] Ram S, Zhang W, Williams M, et al. Predicting asthma-related emergency department visits using big data[J]. IEEE J Biomed Health Inform, 2015, 19(4): 1216-1223. DOI: 10.1109/JBHI.2015.2404829.

[5] Taylor RA, Pare JR, Venkatesh AK, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach[J]. Acad Emerg Med, 2016, 23(3): 269-278. DOI: 10.1111/acem.12876.

[6] 朱志勇,陈一昕,李建功.大数据技术在医疗急重症领域的应用[J].邮电设计技术,2016(8):28-32. DOI: 10.16463/j.cnki.issn1007-3043.2016.08.006.

Zhu ZY, Chen YX, Li JG. Big data application in critical diseases[J]. Des Tech Posts Telecommun, 2016(8): 28-32. DOI: 10.16463/j.cnki.issn1007-3043.2016.08.006.

[7] 贾立静,张恒,赵宇卓,等.大数据视角下的急诊致命性消化道再出血核心指标解析[J].中华危重病急救医学,2018,30(12):1190-1195. DOI: 10.3760/cma.j.issn.2095-4352.2018.12.017.

Jia LJ, Zhang H, Zhao YZ, et al. Analysis of core indicators of fatal gastrointestinal rebleeding in emergency from the perspective of big data[J]. Chin Crit Care Med, 2018, 30(12): 1190-1195. DOI: 10.3760/cma.j.issn.2095-4352.2018.12.017.

[8] 赵宇卓,赵小柯,潘菲,等.急救大数据与临床决策支持[J].中华危重病急救医学,2019,31(1):34-36. DOI: 10.3760/cma.j.issn.2095-4352.2019.01.008.

Zhao YZ, Zhao XK, Pan F, et al. Big data in emergency and clinical decision support system[J]. Chin Crit Care Med, 2019, 31(1): 34-36. DOI: 10.3760/cma.j.issn.2095-4352.2019.01.008.

[9] 陈欣颖,陈愉生,洪如钧,等.老年重症肺炎危险因素及预后分析[J].临床肺科杂志,2011,16(4):499-501. DOI: 10.3969/j.issn.1009-6663.2011.04.003.

Chen WY, Chen YS, Hong RJ, et al. Risk factors and prognostic factors analysis of elderly patients with severe pneumonia[J]. J Clin Pulmonary Med, 2011, 16(4): 499-501. DOI: 10.3969/j.issn.1009-6663.2011.04.003.

[10] 郭丽.186例老年肺炎临床特点和治疗分析[J].临床肺科杂志,2012,17(2):258-259. DOI: 10.3969/j.issn.1009-6663.2012.02.033. Guo L. To explore the clinical characteristics and elderly pneumonia treatment[J]. J Clin Pulmonary Med, 2012, 17(2): 258-259. DOI: 10.3969/j.issn.1009-6663.2012.02.033.

[11] 曾宗鼎,邢崇浩.血清降钙素原及和肽素水平对危重患者医院获得性肺炎的预后评估价值[J].中国急救医学,2017,37(4):340-345. DOI: 10.3969/j.issn.1002-1949.2017.04.013.

Zeng ZD, Xing CH. Value of serum procalcitonin and copeptin in predicting prognosis of hospital-acquired pneumonia in critically ill patients[J]. Chin J Crit Care Med, 2017, 37(4): 340-345. DOI: 10.3969/j.issn.1002-1949.2017.04.013.

(收稿日期:2020-06-11)

关于经过广告审批后的广告中存在不规范医学名词术语未予更改的声明

依照广告审批的相关规定,按照广告厂家的要求,本刊刊登的新活素、血必净及佳维体广告图片和内容均按照广告审查批准文件的原件刊出,故广告内容中“适应症”“禁忌症”未按标准医学名词术语修改为“适应证”“禁忌证”,“其它”未修改为“其他”,“成份”未修改为“成分”,时间单位仍用汉字表示,剂量单位“ml”未修改为“mL”,标示数值范围的标点符号“-”未修改为“~”。特此声明!