

基于 E-CHAID 算法的肺结核患者住院费用 DRG 分组研究

丁志伟 周小平 张海燕

首都医科大学附属北京胸科医院 北京市结核病胸部肿瘤研究所财务处,北京 101149

[摘要] **目的** 探索肺结核患者疾病诊断相关分组(DRG)方案及费用标准。**方法** 收集首都医科大学附属北京胸科医院 2022 年 1 月至 2024 年 12 月 14 756 例肺结核住院患者的临床数据。分析影响患者住院费用的因素;将关键影响因素作为决策树输入变量,运用 E-CHAID 算法生成 DRG 分组,并评估组内同质性和组间异质性。**结果** 不同性别、年龄、婚姻状况、付费方式、住院日、其他诊断数量、是否开展手术、是否危重症、入院途径、离院方式的肺结核患者住院费用比较,差异有统计学意义($P<0.01$)。性别、年龄、住院日、其他诊断数量、是否开展手术、是否危重症、离院方式是肺结核患者住院费用的影响因素($P<0.01$)。E-CHAID 算法生成 10 个 DRG 分组,不同 DRG 分组的肺结核患者住院费用比较,差异有统计学意义($P<0.01$);变异系数均值 0.55。各组内同质性高,组间异质性强,分组合理性良好。**结论** E-CHAID 算法构建的肺结核 DRG 分组方案符合临床实际,住院日、手术、其他诊断和危重症是费用控制的关键靶点。医院需优化非手术患者住院日管理,强化并发症早期筛查,对高权重危重症组实施标准化诊疗路径。建议结合分组权重动态调整医保支付标准,提升资源利用效率。

[关键词] 疾病诊断相关分组;肺结核;E-CHAID 算法;决策树模型;住院费用

[中图分类号] R197.1

[文献标识码] A

[文章编号] 1673-7210(2025)11(c)-0063-05

DOI:10.20047/j.issn1673-7210.2025.33.12

DRG grouping study on hospitalization cost of pulmonary tuberculosis patients based on E-CHAID algorithm

DING Zhiwei ZHOU Xiaoping ZHANG Haiyan

Finance Office, Beijing Chest Hospital, Capital Medical University Beijing Tuberculosis and Thoracic Tumor Research Institute, Beijing 101149, China

[Abstract] **Objective** To explore the diagnosis-related groups (DRG) scheme and cost standards for pulmonary tuberculosis patients. **Methods** The clinical data of 14 756 inpatients with pulmonary tuberculosis in Beijing Chest Hospital, Capital Medical University from January 2022 to December 2024 were collected. The factors influencing patients' hospitalization costs were analyzed; the key influencing factors were taken as the input variables of decision tree, and E-CHAID algorithm was used to generate DRG groups, and the homogeneity within the groups and the heterogeneity between groups were evaluated. **Results** There were statistically significant differences in the hospitalization costs of pulmonary tuberculosis patients of different genders, ages, marital statuses, payment methods, length of hospital stay, the number of other diagnoses, whether surgery was performed, whether they were in critical condition, admission routes, and discharge methods ($P<0.01$). Gender, age, length of hospital stay, the number of other diagnoses, whether surgery was performed, whether they were in critical condition, and discharge methods were the factors affecting the hospitalization costs of patients with pulmonary tuberculosis ($P<0.01$). The E-CHAID algorithm generated 10 DRG groups. The comparison of hospitalization costs among pulmonary tuberculosis patients in different DRG groups showed statistically significant differences ($P<0.01$); the mean coefficient of variation was 0.55. The homogeneity within each group was high, the heterogeneity between groups was strong, and the rationality of grouping was good. **Conclusion** The pulmonary tuberculosis DRG grouping scheme constructed by the E-CHAID algorithm is in line with clinical practice, length of hospital stay,

[基金项目] 北京市卫生经济学会第一批资助立项课题。

[作者简介] 丁志伟(1993.8-),男,硕士;研究方向:医院成本管理。

[通讯作者] 周小平(1983.3-),男,高级会计师,首都医科大学附属北京胸科医院 北京市结核病胸部肿瘤研究所财务处处长;研究方向:医院财务管理。

surgeries, other diagnoses, and critical illnesses are key targets for cost control. Hospitals need to optimize the management of length of stay for non-surgical patients, strengthen early screening for complications, and implement standardized diagnosis and treatment paths for high-weight critical groups. It is recommended to dynamically adjust the medical insurance

payment standards in combination with grouping weights to improve the efficiency of resource utilization.

[Key words] Diagnosis related groups; Pulmonary tuberculosis; E-CHAID algorithm; Decision tree model; Hospitalization cost

结核病是全球第二大单一传染源致死疾病^[1];我国估算结核病发病数排全球第 3 位,2024 年死亡数约 2.5 万,死亡率为 2/10 万,防治压力大^[2];其中肺结核最常见且治疗周期长。疾病诊断相关分组(diagnosis related groups, DRG) 依据国际疾病分类(international classification of diseases, ICD),按照“临床过程一致、资源消耗相似”原则分组并制定费用偿付标准,在提升诊疗质量等方面成效显著^[3]。现有研究虽能优化医保付费、缓解患者经济压力^[4-5];但因地区差异,最终在医保付费优化等方面呈现的效果也有较大差距。故本研究以首都医科大学附属北京胸科医院肺结核患者为例,分析住院费用影响因素并科学分组。

1 资料与方法

1.1 一般资料

本研究数据来源于首都医科大学附属北京胸科医院住院病案首页系统。纳入标准:出院主要诊断在 ICD-10 中编码为 A15(呼吸道结核,经细菌学和组织学证实)、A16(呼吸道结核,未经细菌学或组织学所证实)^[6]。排除标准:①病案首页主要信息存在缺项;②病案首页信息存在逻辑错误;③住院日 > 60 d;④住院费用在 $P_{1\sim}P_{99}$ 范围外^[7]。最终纳入首都医科大学附属北京胸科医院 2022 年 1 月至 2024 年 12 月 14 756 例肺结核住院患者数据。

1.2 研究方法

本研究通过专家咨询、文献查阅与临床实践,采集患者性别、年龄、付费方式、住院日、其他诊断数量、是否手术等病案信息。首先经单因素分析筛选有差异的变量,年龄、住院日参考均值及四分位值,其他诊断数量参考闫晓婧等^[8]分组,性别、婚姻、付费方式、入院途径、离院方式由病案首页系统分组,并参考相关研究分组赋值后,通过多元线性回归分析明确各变量对住院费用的影响程度^[8-11]。E-CHAID 算法设其最大树深度 3 层,父/子分支最小记录数 100/50,拆分节点、合并类别。最后采用 E-CHAID 算法构建 DRG 分组模型。

1.3 统计学方法

采用 SPSS 27 统计学软件进行数据分析。不符合正态分布的计量资料以中位数和四分位数 $[M(P_{25}, P_{75})]$ 表示,两组间比较采用 Mann-Whitney U 检验,多组间比较采用 Kruskal-Wallis H 检验;计数资料以例数表示。将差异有统计学意义的变量作为解释变量,经对数转换后的住院费用作为被解释变量纳入线性回归模型,进一步分析影响住院费用的相关因素。以 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 肺结核住院患者基本情况

14 756 例肺结核患者中男 8 263 例,女 6 493 例;中位年龄 46 岁;中位住院日 6 d;中位住院费用 14 263.49 元。具体情况见表 1。

表 1 不同时间肺结核患者的基本信息 $[M(P_{25}, P_{75})]$

时间	年龄(岁)	住院日(d)	住院费用(元)
2022 年	44(29,61)	8(3,12)	16 859.53(11 042.86,23 860.26)
2023 年	46(30,62)	6(1,10)	14 645.56(8 471.46,22 138.53)
2024 年	48(32,65)	4(1,8)	12 277.62(6 401.26,19 345.10)
平均	46(31,63)	6(1,10)	14 263.49(7 974.99,21 519.24)

2.2 不同特征的肺结核患者住院费用比较

不同性别、年龄、婚姻状况、付费方式、住院日、其他诊断数量、是否开展手术、是否危重症、入院途径、离院方式的肺结核患者住院费用比较,差异有统计学意义($P < 0.01$);不同民族的肺结核患者住院费用比较,差异无统计学意义($P > 0.05$)。见表 2。

2.3 肺结核患者住院费用的影响因素

将表 2 中差异有统计学意义($P < 0.05$)的变量作为解释变量纳入线性回归模型。多重共线性分析结果显示,方差膨胀因子均 < 5 ,容差 > 0.2 ,各变量均不存在共线性。结果显示,性别、年龄、住院日、其他诊断数量、是否开展手术、是否危重症、离院方式是肺结核患者住院费用的影响因素($P < 0.01$)。见表 3~4。

2.4 基于 E-CHAID 算法的 DRG 分组结果

根据表 4 结果构建决策树模型,最终输出 1 个 3 层,17 个节点,10 个终端节点的决策树。7 个变量均纳入决策树,住院日对住院费用影响最为重要,其次是否开展手术、其他诊断数量、是否危重症。不同 DRG 分组的肺结核患者变异系数(coefficient of variation, CV)平均值为 0.55,分组较为合理。见表 5。

2.5 肺结核患者 DRG 分组分析

2.5.1 分组效果多维度评价

2.5.1.1 组间同质性 不同 DRG 分组的肺结核患者住院费用比较,差异有统计学意义($H = 6 760.531, P < 0.001$)。组 6(危重症手术)平均住院费用是组 2(无并发症短住院)的 7.54 倍,占样本 0.40%的组 6 消耗 1.19%资源;组 4~6(开展手术)平均住院费用为 33 941.59 元,是不开展手术的 2.35 倍,凸显手术对费用的驱动作用。

2.5.1.2 组内同质性 各组费用 CV 值范围为 0.22~0.80,呈“重症集中、轻中症分散”;组 6(CV=0.22)、组 5

表 2 不同特征的肺结核患者住院费用比较[元, $M(P_{25}, P_{75})$]

项目	例数	住院费用	Z/H/ χ^2 值	P 值
性别			17.486	<0.001
男	8 263	15 608.65(9 766.39, 22 530.55)		
女	6 493	12 265.58(6 208.21, 19 991.79)		
年龄(岁)			611.544	<0.001
<40	6 199	12 037.96(6 029.25, 18 796.06)		
40~60	4 399	14 614.06(8 359.23, 22 181.39)		
>60	4 158	17 356.63(11 315.70, 23 588.37)		
婚姻状况			185.637	<0.001
未婚	3 378	12 234.08(6 340.42, 19 695.36)		
已婚	10 780	14 697.28(8 419.27, 21 877.19)		
丧偶	371	18 633.69(12 357.14, 26 177.79)		
离异	218	14 001.19(7 960.29, 21 818.52)		
其他	9	11 374.95(6 787.31, 19 555.11)		
民族			1.262	0.207
汉族	14 128	14 294.20(8 058.00, 21 509.54)		
少数民族	628	13 496.65(6 808.23, 21 632.56)		
付费方式			61.845	<0.001
居民医保	12 221	13 996.41(7 845.98, 21 372.83)		
新农合	36	15 975.93(9 041.44, 25 932.35)		
商业保险	17	14 571.41(9 567.10, 21 601.39)		
自费	1 906	15 658.60(10 126.30, 23 086.56)		
其他	576	13 027.70(6 528.29, 19 766.42)		
住院日(d)			5 627.013	<0.001
<8	9 303	10 715.90(5 052.16, 15 328.07)		
8~14	4 144	20 518.47(15 873.27, 26 188.29)		
>14	1 309	29 437.82(21 999.61, 40 661.88)		
其他诊断数量(个)			1 716.824	<0.001
0	355	6 609.94(4 593.12, 8 941.02)		
1~5	6 654	11 583.95(5 668.23, 17 465.49)		
6~10	4 870	16 009.99(10 416.23, 22 239.26)		
11~15	2 032	18 969.43(13 359.78, 25 460.11)		
16~20	824	21 811.78(15 691.57, 30 169.75)		
>20	21	24 032.96(21 264.15, 32 156.20)		
开展手术			10.901	<0.001
否	14 143	14 121.47(8 063.61, 21 120.74)		
是	613	24 090.26(5 586.40, 46 563.37)		
危重症			16.223	<0.001
否	14 625	14 156.05(7 906.85, 21 319.30)		
是	131	46 594.86(32 049.00, 61 995.83)		
入院途径			259.756	<0.001
急诊	1 012	19 276.80(12 775.44, 27 098.01)		
门诊	13 668	13 837.49(7 618.18, 21 090.88)		
其他医疗机构转入	6	16 714.95(4 086.87, 32 015.80)		
其他	70	16 157.29(11 768.83, 21 924.41)		
离院方式			26.780	<0.001
医嘱离院	14 658	14 233.07(7 955.13, 21 466.91)		
医嘱转院	8	22 240.26(14 101.51, 25 545.87)		
非医嘱离院	32	18 744.46(12 335.86, 35 148.01)		
死亡	54	22 724.19(11 848.84, 31 797.85)		
其他	4	19 387.13(12 243.24, 35 344.92)		

(CV=0.40)重症患者标准化诊疗路径执行效果好,费用波动小,轻中症组(2、7、8组)CV>0.55,可能存在漏诊的隐性并发症。

2.5.2 不同分组住院费用分析

2.5.2.1 诊疗强度分析 ①住院日为首要影响因素,每增加 1 d 平均住院费用平均增加 9 959.75 元。住院日

表 3 变量赋值信息

项目	赋值
性别	男=1,女=2
年龄	≥61 岁=0,40~60 岁=1,<40 岁=2
婚姻	未婚=1,已婚=2,丧偶=3,离异=4,其他=9
付费方式	居民医保=1,新农合=2,商业保险=3,自费=4,其他=9
住院日	<8 d=1,8~14 d=2,>14 d=3
其他诊断数量	0 个=1,1~5 个=2,6~10 个=3,11~15 个=4,16~20 个=5,>20 个=6
开展手术	否=0,是=1
危重症	否=0,是=1
入院途径	急诊=1,门诊=2,其他医疗机构转入=3,其他=9
离院方式	医嘱离院=1,医嘱转院=2,非医嘱离院=3,死亡=4,其他=9

表 4 肺结核患者住院费用多元线性回归分析结果

项目	B	S.E.	β	t 值	P 值	容差	VIF
性别	-797.975	154.609	-0.033	5.161	<0.001	0.944	1.060
年龄	-546.941	112.492	-0.038	4.862	<0.001	0.643	1.554
婚姻	27.736	148.258	0.001	0.187	0.852	0.805	1.242
付费方式	-31.435	42.031	-0.005	0.748	0.455	0.992	1.008
住院日	9 959.753	124.994	0.546	79.682	<0.001	0.836	1.197
其他诊断数量	1 232.853	96.707	0.096	12.748	<0.001	0.690	1.448
开展手术	9 315.486	384.955	0.156	24.199	<0.001	0.942	1.062
危重症	18 985.829	827.669	0.150	22.939	<0.001	0.922	1.084
入院途径	92.042	137.383	0.004	0.670	0.503	0.981	1.019
离院方式	695.46	247.049	0.018	2.815	0.005	0.955	1.048

注 VIF:方差膨胀因子。

表 5 肺结核患者 DRG 分组结果

分组	组合名称	例数	平均住院费用(元)	CV
组 1	住院日 ≤7 d,其他诊断数量 11~21 个	972	14 142.88	0.55
组 2	住院日 ≤7 d,无合并并发症	327	7 369.29	0.80
组 3	住院日 8~14 d,未开展手术	3 927	21 729.32	0.44
组 4	住院日 ≥15 d,开展手术	1 224	31 464.28	0.43
组 5	住院日 ≥8 d,开展手术,非危重症	250	41 573.64	0.40
组 6	住院日 ≥8 d,开展手术,危重症	52	55 561.28	0.22
组 7	住院日 ≤7 d,其他诊断数量 1~5 个,女性	2 961	9 775.71	0.72
组 8	住院日 ≤7 d,其他诊断数量 1~5 个,男性	2 269	11 304.25	0.71
组 9	住院日 ≤7 d,其他诊断数量 6~10 个或者 ≥21 个,女性	1 212	11 817.84	0.67
组 10	住院日 ≤7 d,其他诊断数量 6~10 个或者 ≥21 个,男性	1 562	13 019.65	0.58

注 DRG:疾病诊断相关分组;CV:变异系数。

从 ≤7 d 延长至 8~14 d,平均住院费用增长约 96.52%。从 8~14 d 延长至 ≥15 d,平均住院费用增长约 43.47%。②手术操作因素。手术患者住院费用比非手术患者高 9 315.49 元,其中组 6 平均住院费用是组 5 的 1.34 倍,“手术 + 重症”对平均住院费用有叠加效应。③并发症数量每增加 1 个,平均住院费用增加 1 232.85 元,并发症 ≥11 个的患者平均住院费用(20 180.40 元)是无并发症患者(6 609.94 元)的 3.05 倍。

2.5.2.2 病情严重程度分析 危重症患者平均住院费用比非危重症高 18 985.83 元,且住院日显著增加,反映重症患者需更多医疗资源与时间支持。

2.5.2.3 人口学因素分析 女性患者平均住院费用比男性患者低 797.98 元,性别差异明显,或因男性外出务

工多、工作环境差、劳动强度大、作息不规律等导致结核菌感染及发病机会更高。

3 讨论与建议

3.1 分组的科学性与有效性

采用 E-CHAID 算法对肺结核患者住院费用进行 DRG 分组,能处理医疗数据中的大量分类变量,且能构建可视化规则树辅助临床决策。通过与已有研究对比,本研究既验证了住院日、手术操作、并发症数量等对住院费用的影响,还借助精细的分组进一步揭示了不同疾病程度患者的住院费用差异,体现分组方法的科学性与有效性。

3.2 肺结核患者管理的应用建议与策略优化

首先,针对非手术患者,尤其是住院日较长的患

者,优化住院日管理,设置费用预警线,可避免无效住院日导致的费用超支。其次,对于危重症手术患者,应严格执行标准化诊疗路径,确保资源的高效利用。同时,医院还应加强并发症的筛查和诊断,减少因漏诊导致的费用增加。此外,医院应充分利用数据挖掘和机器学习技术,对病种数据深入分析,提供管理决策依据。

3.3 医院病种管理路径优化建议与对策

在 DRG 支付下,医院需针对肺结核制订科学的管理路径,以优化资源配置、提升诊疗效率和质量。对医院病种管理路径优化建议如下:

加强病种数据分析与利用。医院应充分利用病案首页、电子病历系统等数据源,对特定病种的患者特征、诊疗过程、费用结构等进行深入分析。通过数据挖掘和机器学习等技术手段,发现潜在的诊疗规律和费用控制点,为管理决策提供支持^[12]。

优化诊疗路径与流程。基于病种数据分析结果,医院应针对特定病种制订或优化诊疗路径和流程。通过减少不必要的检查和治疗项目、缩短住院日等措施,降低患者医疗费用,提升医院运营效率^[13]。

加强医疗质量管理与控制。医院应建立健全医疗质量管理体系,加强对特定病种诊疗过程的监控和评估。通过定期的质量检查、病例讨论、多学科会诊等方式,及时发现并纠正诊疗过程中的问题,确保医疗质量和安全。

强化成本意识与费用控制。医院需强化成本意识,加强对医疗费用的控制。通过建立费用预警机制、实施成本核算与绩效考核等措施,引导医务人员合理使用医疗资源,降低不必要的费用支出。

加强患者教育与沟通。医院应加强对患者的健康教育和沟通工作,提高患者对特定病种诊疗过程、费用结构等方面的认知水平。通过增强患者的自我管理能力和遵医行为,降低医疗费用负担。

4 小结

基于 E-CHAID 算法的肺结核患者住院费用 DRG 分组研究为医院病种管理提供了科学依据和有益参考。通过加强病种数据分析与利用、优化诊疗路径与流程、加强医疗质量管理与控制、强化成本意识与费用控制,以及加强患者教育与沟通等措施的实

施,医院可以进一步提升病种管理水平,降低患者医疗费用负担,提升整体运营效率和质量。

利益冲突声明:本文所有作者均声明不存在利益冲突。

[参考文献]

- [1] 刘星媛,王敏,左秀然,等. 59299 例肺结核患者住院费用结构分析[J]. 中国病案,2025,26(4):42-45.
- [2] 舒薇,刘宇红. 矢志革新履践致远:《2024 年全球结核病报告》结核病科学研究章节解读[J]. 中国防痨杂志,2025,47(2):137-141.
- [3] 高辰旭,冯文. DRG 改革对公立医院住院服务效率及质量的影响[J]. 卫生经济研究,2023,40(12):42-45.
- [4] 闫晓婧,聂文娟,刘海燕,等. 肺结核患者住院费用影响因素与按疾病诊断相关分组方案研究[J]. 中国防痨杂志,2023,45(11):1084-1089.
- [5] 胡鹏,吴成璧. 耐药肺结核患者治疗结局及费用补偿影响因素分析[J]. 卫生经济研究,2024,41(8):41-44.
- [6] 李改云,吴杨昊天,丁明峰,等. 取消药品加成政策前后肺结核患者住院费用的中断时间序列分析[J]. 中国防痨杂志,2022,44(10):1071-1078.
- [7] 程广辉,闫书铭,时松和,等. 神经系统疾病诊断相关分组研究——以河南省某三甲医院为例[J]. 现代预防医学,2021,48(10):1912-1916.
- [8] 周鹏飞,杨孝光,路强,等. 恶性肿瘤术后化疗住院患者 DRG 分组研究——基于 E-CHAID 算法[J]. 卫生经济研究,2023,40(6):35-39,43.
- [9] 张芬,余金明,陈秋艳. Exhaustive CHAID 分类树与 logistic 回归在脑卒中危险因素中的应用[J]. 中国预防医学杂志,2011,12(7):573-576.
- [10] 邵慧丽,宁传英. 基于 E-CHAID 算法胆囊结石患者 DRGs 分组研究[J]. 中国卫生统计,2019,36(4):554-555,559.
- [11] 苏勇. 基于决策树 E-CHAID 算法建立股骨颈骨折病例的 DRGs 分组[J]. 中国病案,2022,23(10):62-66.
- [12] 金萍妹,华伟,陈洁,等. 基于疾病诊断相关组法制定单病种住院费用标准的研究[J]. 中国卫生经济,2017,36(2):26-28.
- [13] 张娴静,王爱荣. DRG 相对权重计算方法比较与探讨[J]. 中国医院管理,2020,40(9):70-71,76.

(收稿日期:2025-06-10)

(修回日期:2025-09-16)