

基于决策树模型的早产儿病例住院费用的分组研究*

彭伊婧^{①②}, 孙 琰^{①②}, 余仁强^②, 蒋 频^②, 周倩雯^②, 钱俊益^②

摘要 目的: 研究早产儿住院患者的病例组合方式及费用支付标准, 为制定住院费用分组以及提高医疗效率提供参考依据。方法: 收集样本三级妇幼专科医院2019—2021年3 061例早产儿病例, 采用多元线性逐步回归分析, 确定住院费用的主要影响因素, 运用决策树模型中CHAID方法构建早产儿患者住院费用分组方案及付费标准。结果: 住院天数、合并症或并发症个数、出生体重、有无操作、使用呼吸机时长及转归情况作为影响早产儿病例住院费用的分类节点, 利用决策树模型对3 061例早产儿患者分组, 形成7个DRG组合并对应住院费用标准。结论: 早产儿病例DRG分组能够客观反映医疗资源消耗水平, 可为其他地区医疗机构分组进行验证并优化分组方案提供参考依据。

关键词 疾病诊断相关分组; 早产儿; 住院费用; 决策树模型

中图分类号 R1-9; F840.613 **文献标志码** A **文章编号** 1003-0743(2023)04-0030-04

Research on Grouping Hospitalization Costs of Patients with Premature Infants Based on Decision Tree/PENG Yi-jing, SUN Yan, YU Ren-qiang, et al./Chinese Health Economics, 2023,42(4):30-33

Abstract Objective: To study the case mix mode and cost payment standard of patients with premature infants, and to provide a reference for formulating hospitalization expenses groups and improving medical efficiency. **Methods:** A total of 3 061 cases of premature infants in a tertiary maternity and children's hospital from 2019 to 2021 were collected. Multiple linear stepwise regression analysis was used to determine the main influencing factors of hospitalization expenses. The CHAID method in decision tree model was used to construct the hospitalization expense grouping plan and payment standard for premature infants. **Results:** The number of days in hospital, number of comorbidities or complications, neonatal birth weight, presence or absence of operations, length of time on ventilator, and transfer status were used as classification nodes to influence the hospitalization cost of preterm cases, and a total of 7 DRG combinations and corresponding hospitalization cost criteria were formed using a decision tree model to group 3 061 preterm patients. **Conclusion:** Case-mix diagnosis related groups of premature infants could objectively reflect the level of medical resource consumption, and provide a reference for medical institutions in other regions to verify the grouping and optimize the grouping plan.

Keywords decision tree; Diagnosis Related Groups; premature infant; hospitalization cost

First-author's address Medical Quality Management Division, Wuxi Maternity and Child Health Care Hospital, Wuxi, Jiangsu, 214000, China

Corresponding author SUN Yan, E-mail: 2423193934@qq.com

据世界卫生组织的调查资料^[1]显示, 全球早产儿的发生率超过10%, 而我国早产儿发生率约为11%, 且有逐年上升趋势。研究表明^[2-4], 早产儿胎龄小、体质量轻、各器官及组织发育不完善, 约有25%以上的早产儿有可能受到神经发育问题的影响, 部分患儿会出现严重的神经系统后遗症, 如认知、运动和语言障碍等问题, 影响早产儿的生命质量, 加剧了医疗资源的投入, 产生高昂的医疗费用, 为社会和许多家庭带来难以负担的经济压力。因此, 研究早产儿病例住院费用的分组在医疗支付改革和医疗保险费用控制方面具有重要意义。疾病诊断相关分组是将疾病特征相似、诊疗过程相近、资源消耗程度近似的病例分入同一

DRG组中^[5]。一方面能够反映病例的实际诊疗需求和医疗服务利用情况, 另一方面为医保支付、临床质量控制及医疗服务绩效考核与评价提供重要的参考依据。本研究采用决策树模型为早产儿病例建立DRG分组, 为本地区早产儿病例的适配分组和合理制定医保支付标准提供相关依据。

1 资料与方法

1.1 资料来源

本研究从无锡市样本三甲妇幼专科医院病案首页提取相关研究数据。纳入标准: (1) 疾病诊断编码包括P07.2(极度不成熟儿)和P07.3(其他早产婴儿); (2) 患儿出院科室为新生儿科; (3) 出院时间为2019年1月1日—2021年12月31日。排除标准: (1) 病案首页存在重要数据缺失、不符合逻辑及填写错误; (2) 住院总费用与各类费用相加不一致; (3) 住院天数小于1天。最终收集到有效的早产儿病例3 061例。

1.2 方法

收集早产儿病例的性别、出生体重、住院天数、合并症或并发症个数、有无操作、使用呼吸机时长及转归等住院病案首页数据并进行整理, 详见表1。

* 基金项目: 江苏省妇幼保健重点学科项目(SFY-EK2021); 无锡市医学创新团队项目(CXTD2021013)。

① 江南大学附属儿童医院(无锡市儿童医院)

江苏 无锡 214000

② 江南大学附属妇产医院

江苏 无锡 214000

作者简介: 彭伊婧(1994—), 女, 硕士学位, 助理研究员;

研究方向: 医保支付与改革; E-mail: 727182031@qq.com。

通信作者: 孙琰, E-mail: 2423193934@qq.com。

表1 早产儿病例基本特征及赋值

变量	变量名称	赋值说明
X ₁	性别	男=1, 女=2
X ₂	住院天数	≤9天=1, 10~16天=2, >16天=3
X ₃	出生体重	<1 kg=1, 1~1.5 kg=2, 1.5~2.5 kg=3, >2.5 kg=4
X ₄	合并症或并发症个数	<4个=1, ≥4个=2
X ₅	有无操作	无=1, 有=2
X ₆	使用呼吸机的时长	无=0, <96 h=1, >96 h=2
X ₇	转归	死亡/自动出院=1, 治愈/好转=2

1.3 统计学方法

本研究采用SPSS 24.0软件进行统计分析, 利用多元线性逐步回归法筛选出重要的影响因素并作为决策树模型的分层节点, 采用决策树模型中CHAID方法构建DRG分组方案, 计算各个分组的住院费用标准、费用上限。决策树的参数设置: 父节点最小病例数为300, 子节点最小病例数为150, 设置树的最大深度为3层, 如果节点上的例数达不到所设置的要求, 那么该节点就是终点, 拆分和组合显著性检验水准 $\alpha=0.05$ 。DRG分组合理性评价方法: 采用非参数检验法检验组间异质性, 计算变异系数(CV)检验组内异质性^[6]。

2 结果

2.1 一般情况

本研究有效样本共计3 061例。其中, 男性1 701例, 占55.57%, 女性1 360例, 占44.43%。出生体重<1 kg的26例, 占0.85%; 1~1.5 kg的236例, 占7.71%; >1.5~2.5 kg的1 825例, 占59.62%; >2.5 kg

的974例, 占31.82%; 平均住院天数13.81天, 中位数9天; 平均住院费用13 714.33元, 中位数8 079.73元。

2.2 早产儿病例住院费用影响因素

K-S正态性检验显示, 早产儿病例的住院费用不符合正态分布, 呈现偏态分布。因此, 对其进行对数转换, 将对数转换后的住院费用作为因变量, 以性别、出生体重、住院天数、使用呼吸机时长、有无操作、合并症或并发症个数、转归作为自变量, 构建出多元线性回归模型, 确定影响因素。结果显示, 住院天数、出生体重、合并症或并发症个数、有无操作、使用呼吸机时长、转归6个变量是主要影响因素。模型调整后的决定系数 $R^2=0.749$, 表明74.90%的住院费用变动是受到上述6个因素的影响产生的。方差膨胀因子均<5, 可认为各变量间没有明显的多重共线性, 表明该多元线性回归模型拟合良好, 详见表2。

2.3 早产儿决策树模型病例分组预测

运用多元线性回归分析筛选出有统计学意义的影响因素作为决策树的分类节点。结果显示: 住院天数、出生体重、有无操作、合并症或并发症个数4个变量进入决策树模型, 模型分为3层, 共形成7个组合, 详见图1。模型风险为0.035, 各组合的变异系数CV均<1, 此结果表明各个组内住院费用的同质性高, 变异小; 各DRG组合经过非参数假设检验, 显示7个病例组合住院费用分布不全相同, 组间差异有统计学意义($H=22\ 868.560, P<0.001$)。综合分析各项指标, 该分组方案较为理想, 详见表3。

2.4 住院费用病例组合标准及病种权重

标准费用设定为各组合早产儿病例住院费用的中

表2 早产儿住院费用多元线性回归分析结果

因素	b	SE	b'	t	P	b的95% CI	容差	方差膨胀因子
住院天数分类	0.257	0.005	0.590	52.805	0	0.248~0.267	0.656	1.525
有无操作	0.193	0.012	0.198	15.548	0	0.169~0.217	0.505	1.981
个数分类	0.133	0.008	0.184	17.534	0	0.118~0.148	0.742	1.349
体重分类	-0.066	0.006	-0.113	-11.295	0	-0.078~-0.055	0.820	1.219
离院方式	0.118	0.033	0.033	3.614	0	0.054~0.182	0.976	1.024
使用呼吸机的时长	0.029	0.013	0.025	2.158	0.031	0.003~0.055	0.620	1.614
常数	3.275	0.067	—	48.892	0	3.143~3.406	—	—

表3 早产儿决策树病例组合方案

组别	病例组合名称	样本量(例)	占比(%)	均数(元)	标准差(元)	CV
DRG ₁	住院天数≤9天, 合并症或并发症的个数<4个	1 124	36.70	5 153.20	1 865.90	0.36
DRG ₂	住院天数≤9天, 合并症或并发症的个数≥4个, 新生儿体重≤2.5 kg	279	9.10	8 802.41	7 222.25	0.82
DRG ₃	住院天数≤9天, 合并症或并发症的个数≥4个, 新生儿体重>2.5 kg	198	6.50	6 681.20	2 384.71	0.36
DRG ₄	住院天数10~16天, 合并症或并发症的个数<4个	344	11.20	9 425.02	2 814.92	0.30
DRG ₅	住院天数10~16天, 合并症或并发症的个数≥4个	369	12.10	12 127.82	3 889.76	0.32
DRG ₆	住院天数>16天, 无操作	412	13.50	23 567.15	12 019.70	0.51
DRG ₇	住院天数>16天, 有操作	335	10.90	44 721.13	19 839.37	0.44

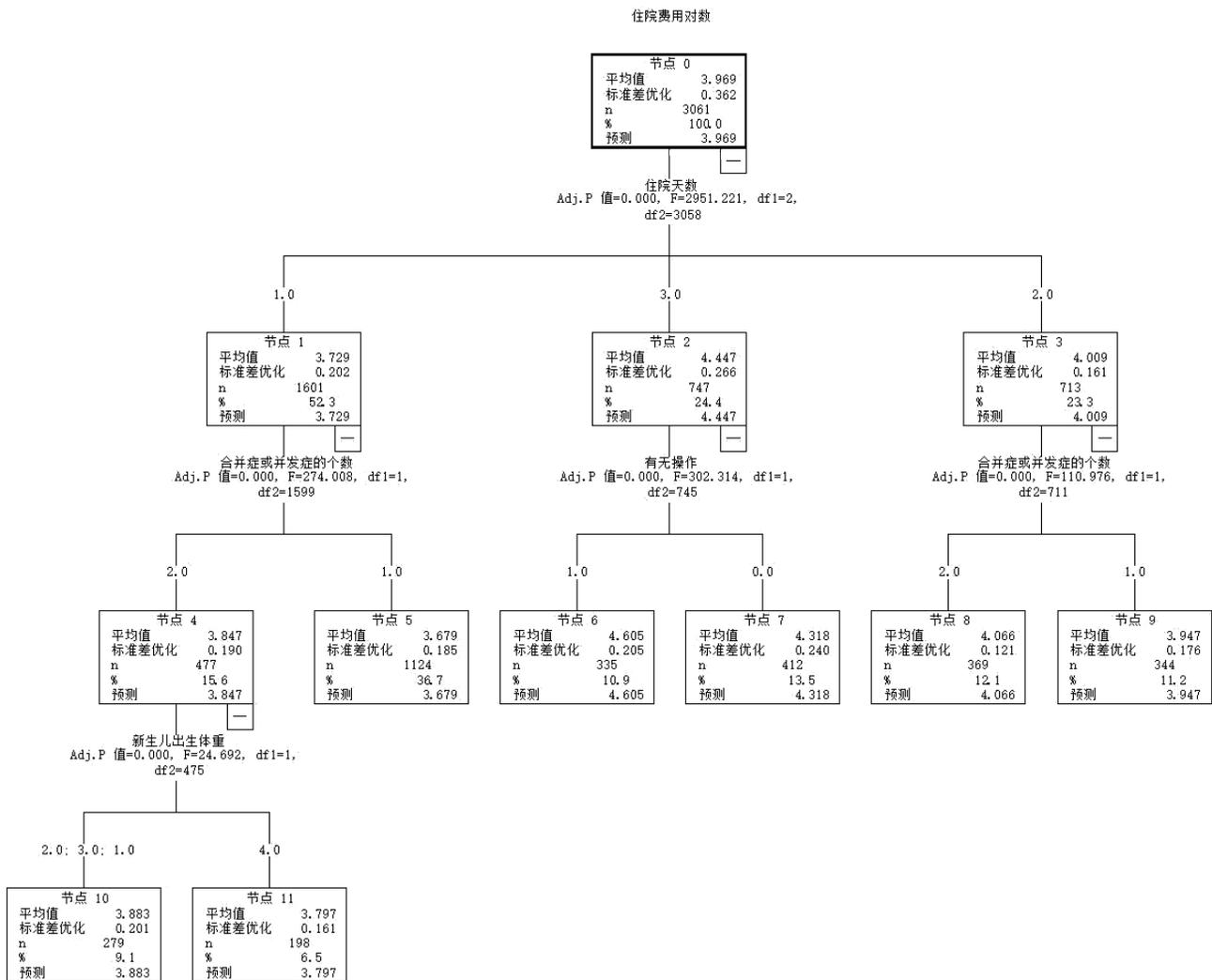


图1 早产儿决策树模型

位数，住院费用的上限设定为各组合住院费用的 P_{75} 加上 1.5 倍四分位间距^[7-9]。各组合中住院费用高于费用上限的病例即被认为超限病例。病种权重为某组合所有病例的平均费用÷所有组合早产儿病例的平均费用，反映了各组合资源消耗的相对程度，权重越高代表消耗的医疗资源越多^[10-11]，详见表4。

3 讨论

DRG 相比于单病种付费，考虑了疾病的复杂性和

患者的特异性，通过制定并及时调整各病组的支付标准，兼顾了各方利益，是目前国际上较为主流、科学的支付方式^[12]。DRG 不仅是医疗保险的支付方式，他还更有利于医务人员根据疾病分组合理制定诊疗方案，减轻患者及家属的疾病经济负担，并且可以作为管理工具应用于医院的经营决策，达到提高资源利用效率，节约成本，提升医院运营效率的目的^[13-14]。决策树模型分组要求各组间差异大、组内差异小，这与 DRG 分组

表4 早产儿各DRG分组住院费用标准

组别	例数 (例)	住院费用 P_{75} (元)	四分位数 间距 (元)	标准费用 (元)	费用上限 (元)	超限 例数 (例)	超限 比例 (%)	病种权重
DRG ₁	1 124	6 238.46	2 992.74	4 985.21	10 727.57	5	0.44	0.38
DRG ₂	279	9 080.40	3 067.01	7 355.70	13 680.92	21	7.53	0.64
DRG ₃	198	8 024.53	2 985.81	6 254.61	12 503.25	4	2.02	0.49
DRG ₄	344	11 107.79	3 530.09	9 390.18	16 402.93	3	0.87	0.69
DRG ₅	369	13 552.12	4 304.10	11 403.73	20 008.27	15	4.07	0.88
DRG ₆	412	28 393.24	13 494.82	20 649.18	48 635.47	21	5.10	1.72
DRG ₇	335	59 991.43	32 374.29	42 007.79	108 552.87	1	0.30	3.26

原理相一致^[15-16]。因此,本研究以早产儿为例进行DRG研究,可探索并制定适合本地早产儿病例的住院费用标准,同时为DRG分组的本土化提供科学的依据。

本研究决策树模型预测结果显示病例分组效果较好,组内同质性高,组间差异性大。分类节点变量将全部病例分成7个DRG组合。第1层分类节点是住院天数,第2层分类节点是合并症或并发症个数、有无操作,第3层分类节点是出生体重。组合1的住院费用和病种权重最低,说明在住院天数 ≤ 9 天时,合并症或并发症个数是影响住院费用的主要因素,合并症或并发症一定程度上可以反映出病情的复杂程度,数量的增加会导致住院时间延长,治疗中所需的医疗资源支出也越大。组合7的住院费用和病种权重最高,说明在住院天数 > 16 天且有操作时,早产儿的治疗难度较大,住院费用较高。提示该病例组合的患儿在住院期间有使用有创呼吸机通气、气管内插管、腰椎穿刺及胸腔穿刺等操作,会增加医疗资源的消耗,导致住院费用的升高,这与张俊清等^[17]的研究结果一致。在住院天数 ≤ 9 天且合并症或并发症个数 ≥ 4 个时,低出生体重儿的病情严重程度和住院费用均比正常出生体重儿高,与低出生体重儿预后较差有关,如何有效提高低出生体重儿救治效率成为降低住院费用的重要途径。

本研究采用中位数为各组合的标准费用, P_{75} 加1.5倍四分位间距为上限费用,考虑到某些特殊诊疗情况,允许各组合中有一定比例的患儿住院费用超过上限,一般认为这应该限制在5%以内^[18],如果费用超过上限的病例比例超出这个范围,则提示医务人员可能存在过度诊疗的情况。组合2、6的超限比例超过5%,尤以组合2(住院天数 ≤ 9 天,合并症或并发症个数 ≥ 4 个,出生体重 ≤ 2.5 kg)最为明显,超限病例比例达7.53%,提示该病例组中部分患儿消耗的医疗资源较多。因此,需重点加强管控费用超过上限的病例比例较高的组合,以规范诊疗行为,有效利用医疗资源,控制费用的不合理增长,实现医保精细化管理。

以往的研究以早产儿的流行病学特点、治疗效果、转归及预后的临床相关研究为主,对影响早产儿住院费用的研究不多,本研究在查询国内外已有文献的基础上,结合了患儿的合并症或并发症情况、出生体重、有无操作及呼吸机使用情况等相关因素对住院费用进行分析,反映了早产儿病例的诊疗情况和实际医疗需求。对早产儿病例进行分组并计算病种权重、标准费用及费用上限等,对医疗机构制定合理的控费措施、科学的绩效评价及临床质量控制标准,实现医疗精细化管理有重要意义,同时也有利于医保部门合理设定该病种不同分组的补偿标准。

参 考 文 献

[1] CHEVALLIER M, DEBILLON T, PIERRAT V, et al. Lead-

- ing causes of preterm delivery as risk factors for intraventricular hemorrhage in very preterm infants: results of the EPIPAGE 2 cohort study[J]. *Obstetric anesthesia digest*, 2018, 216(5):518.e1-518.e12.
- [2] 闫淑媛,刘震宇,钱红艳,等.不同胎龄及出生体质量早产儿早期神经发育的纵向研究[J].*临床儿科杂志*, 2017,35(6):425-429.
- [3] 王晓鹏,田秀英,郑军,等.天津市早产儿流行病学调查[J].*中国妇幼保健*, 2016,31(24):5473-5475.
- [4] SUMANASENA S P, VIPULAGUNA D V, MENDIS M, et al. Beyond survival: 5-year neurodevelopmental follow-up of a cohort of preterm infants in Colombo, Sri Lanka[J]. *Paediatrics and international child health*, 2018,38(2):128-136.
- [5] 董乾,陈金彪,陈虎,等.DRGs国内发展现状及政策建议[J].*中国卫生质量管理*, 2018, 25(2):1-4.
- [6] 韩栋,魏凌云,章莹,等.妇婴疾病诊断相关组的构建与比较研究[J].*中国卫生统计*, 2016, 33(6):986-988.
- [7] 许志隆,刘宝.某医院2型糖尿病患者住院费用的疾病诊断相关分组研究[J].*中国卫生资源*, 2020,23(4):373-377, 399.
- [8] 黄利娟,查君敬,梁学柱,等.2型糖尿病患者住院费用的DRGs分组研究[J].*中国病案*, 2021,22(8):21-24.
- [9] 仲亚琴,肖静,许琼琼,等.宫颈癌患者住院费用的DRGs分组研究[J].*中华医院管理杂志*, 2017,33(10):738-741.
- [10] 王彤彤,周学健,唐琴,等.医院病例组合指数(CMI)评价管理研究[J].*中国病案*, 2021,22(8):24-27.
- [11] 宁传英,邵慧丽.基于决策树模型急性阑尾炎患者DRGs分组研究[J].*中国卫生统计*, 2018,35(4):584-586.
- [12] 王珊,饶克勤,刘丽华.我国按病种付费实施现状及国际DRGs支付借鉴[J].*中华医院管理杂志*, 2015,31(2):81-83.
- [13] 殷玉华,吴华余,朱健倩.基于决策树的胆总管结石患者住院费用分组研究[J].*中国医院统计*, 2021,28(4):321-324.
- [14] 李晓慧,魏峰,邢花.按疾病诊断相关分组付费实施效果——基于沈阳市某三甲医院[J].*现代医院*, 2022,22(3):414-416.
- [15] WU S W, PAN Q, CHEN T. Research on diagnosis-related group grouping of inpatient medical expenditure in colorectal cancer patients based on a decision tree model[J]. *World j clin cases*, 2020,8(12):2484-2493.
- [16] 李娟,金雯,李盈.基于决策树的中耳疾患住院费用分组研究[J].*现代医院*, 2022,22(5):703-705.
- [17] 张俊清,郭亮,李法军,等.新生儿呼吸窘迫综合征患者住院费用的DRGs分组研究[J].*现代预防医学*, 2020,47(19):3513-3517.
- [18] 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.

[收稿日期:2023-01-16] (编辑:杨威)