基于决策树模型的青光眼住院费用 DRG 分组实证研究*

袁筱祺[1] 董 笑[2] 赵 爽[1]

([1] 上海市第一人民医院医务处 上海 200080 [2] 上海市第一人民医院肿瘤科 上海 200080)

〔摘要〕 **目的/意义** 探讨上海市青光眼患者的 DRG 分组和相应住院费用标准，为医疗机构合理控制住院费用提供方法学依据。**方法/过程** 提取 2022 年 1 月 1 日—2023 年 12 月 31 日主要诊断为青光眼的住院患者相关数据，通过单因素分析和多元线性回归初筛分类节点变量。运用数据挖掘技术中决策树模型探讨青光眼住院患者的 DRG 分组方案、住院费用标准及病种权重，利用变异系数、方差膨胀因子、秩和检验测试模型效能。**结果/结论** 基于决策树模型建立的青光眼患者 DRG 分组具有较高的组内同质性，有无并发症、出院情况作为分类节点变量，病例组合的变异系数为 0.33～0.75，均小于 1。最终形成 4 个住院费用标准和病种权重方案，整体较为合理，为医疗资源合理配置提供参考依据。

〔关键词〕 决策树模型；青光眼；住院费用；DRG

〔中图分类号〕 R－058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673－6036.2024.12.008

An Empirical Study on DRG Grouping of Glaucoma Hospitalization Costs Based on Decision Tree Model

YUAN Xiaoqi[1], *DONG Xiao*[2], *ZHAO Shuang*[1]

[1]*Department of Medical Affairs, Shanghai General Hospital, Shanghai 200080, China*; [2]*Department of Oncology, Shanghai General Hospital, Shanghai 200080, China*

〔**Abstract**〕 **Purpose/Significance** To discuss the DRG grouping and the corresponding hospitalization cost standard of glaucoma patients in Shanghai, and to provide methodological basis for the reasonable control of hospitalization costs in medical institutions. **Method/Process** Data of inpatients with the principal diagnosis of glaucoma from January 1, 2022 to December 31, 2023 are extracted, and the classification node variables are classified by single factor analysis and multiple linear regression. The DRG combination scheme, hospitalization cost standard and disease weight of glaucoma inpatients are investigated by using decision tree model in data mining technology. Coefficient of variation, variance inflation factor and rank sum test are used to test the efficacy of the model. **Result/Conclusion** The DRG grouping of glaucoma patients established based on decision tree model has high intra－group homogeneity, and the presence or absence of complications and discharge status are entered into the model as classification node variables. The coefficient of variation of the case combinations is 0.33 to 0.75, all less than 1. Finally, a total of 4 hospitalization cost standards and disease weight schemes are formed, and the overall scheme is reasonable, providing references for rational allocation of medical resources.

〔修回日期〕 2024－06－06

〔作者简介〕 袁筱祺，研究实习员，发表论文 10 余篇；通信作者：赵爽。

〔基金项目〕 国家自然科学基金青年项目（项目编号：82203742）；上海申康医院发展中心市级医院诊疗技术推广及优化管理项目（项目编号：SHDC12022622）；上海交通大学中国医院发展研究院医院管理建设项目（项目编号：CHDI－2019－B－14）。

〔**Keywords**〕 decision tree model; glaucoma; hospitalization costs; DRG

1 引言

青光眼是一组以视神经凹陷性萎缩和视野缺陷为特征的疾病，是导致失明的第 2 大病因，仅次于白内障，但却是第 1 位不可逆的致盲眼病。疾病诊断相关分组（diagnosis - related groups，DRG）是衡量医疗服务质量和效率，以及确定医保支付的重要工具[1-2]，也是评价公立医院高质量发展的关键指标。实际推广应用过程中，医疗行为本身的复杂性和差异性使 DRG 分组方案具有局限性，尤其是新诊疗技术的应用尚不能精准评估医疗资源消耗，导致医疗机构诊疗技术更新积极性降低、患者无法得到最优治疗方案、分次实施手术等负面情况发生。目前国内外各地在采用方差分析、独立样本 t 检验、倾向性匹配等统计学方法的同时融入神经网络、灰色预测、大语言模型等大数据分析方法进行相关实证试点研究。近年来机器学习方法广泛应用于医疗数据处理，非线性对应能力和稳定性均较好，解决了传统数理统计模型中各变量间多重共线性的问题。Liu Y 等[3]比较 10 个国家原始 DRG 版本的异同，并以缺血性卒中为例，确定各种 DRG 系统的设计和实施策略。Li Q 等[4]选择慢性阻塞性肺疾病、急性心肌梗死和脑梗死患者，将常规付费组和 DRG 付费组患者分别作为对照组和干预组，进行倾向得分匹配。Wang H 等[5]引入大语言模型，通过微调临床记录增强 DRG 分配。马原等[6]利用计算数列序列、关联系数等方法探究成都市某三甲医院 DRG 付费下急性白血病患儿住院费用与内部各因素之间的关联。张英强等[7]利用 χ^2 检验分析南京市某老年病医院 DRG 付费改革前后医疗数据的变化，优化老年病医院的管理策略。葛龙等[8]通过变异系数、总体方差减少系数、χ^2 检验等方法，分析 DRG 管理下恶性肿瘤药物治疗基本情况与住院费用影响因素。王文君等[9]采用灰色关联模型对比研究高倍率病例和正常倍率病例患者住院总费用的内部影响因素。上述学者通过 χ^2 检验、变异系数等多种传统统计学方法，关注缺血性卒中、慢性阻塞性肺疾病、急性白血病等多种疾病住院费用的 DRG 分组细则，为国内外医疗资源的合理利用、医院管理的策略调整提供了指导性建议。基于此，本研究以 DRG 改革试点医院的青光眼患者为研究对象，运用单因素及多因素回归分析探索青光眼患者住院费用的影响因素。通过搭建决策树模型进一步对青光眼患者进行 DRG 分组方案细则的探索，以期形成合理科学的青光眼患者医疗费用分组标准，为医疗资源合理配置提供参考依据，为进一步推广应用 DRG 付费方式及成本管理提供方法学依据。

2 资料来源与方法

2.1 资料来源

数据来源于上海市第一人民医院出院日期为 2021 年 1 月 1 日—2023 年 12 月 31 日出院患者中以青光眼为主要诊断（ICD 编码大致包含：H40、E11.3、E14.3、H44.5、Q15 等）的病案首页数据，选取有效字段，共 2 150 例。剔除不满足条件、内容缺失、格式内容不符合、住院天数小于 1 天或大于 60 天的患者，共纳入 2 137 例患者。

2.2 变量赋值

将数据中住院患者住院费用作为因变量，经与临床医生、医保人员交流与查阅相关资料，将可能影响住院费用的字段：性别、年龄、出院情况、是否有并发症、医保类型、住院天数作为自变量。

2.3 研究方法

采用 SPSS 27.0 描述与分析病例数据，定量资料采用 t 检验和 ANOVA 方差分析进行差异性分析，定性资料采用 χ^2 检验构成差异分析。以 $\alpha = 0.05$ 为统计学检验水准。定量变量用“平均数 ± 标准差”描述，定性变量采用“例数/构成比”描述。采用

多元线性逐步回归方法[10]建立回归模型，确定决策树的分类节点变量，标准化系数越大说明该自变量对因变量的影响越大；用方差膨胀因子[11]（variance inflation factor，VIF）进行共线性诊断，若 VIF >10，说明模型中有很强的共线性。采用决策树模型对纳入患者病例分组，并统计各组别住院费用和医疗资源消耗情况。用变异系数（coefficient of variation，CV）表示组内变异情况，CV 值越小，组内变异越小[12]。通过计算 DRG 分组的病情权重，比较各组消耗的医疗资源，各组合的病种权重 = 该组合的平均住院费用/所有病例的平均住院费用，权重越高，消耗的医疗资源越多。

3 结果

3.1 基本情况

共纳入有效病例 2 137 例，男女病例数之比为 0.75∶1；年龄中位数为 68 岁，其中大于 60 岁 1 649 例，在所有年龄段中占比最高；住院天数中位数为 2 天，其中小于 2 天 1 993 例，占比 93.3%；1 652 例患者有并发症，占比 77.3%；医疗付费方式中城镇职工医保占比最高，为 84%。根据单因素分析结果，年龄、性别、出院情况、住院天数、是否有并发症、医保类型均为住院总费用影响因素，差异有统计学意义（$P<0.05$），见表 1。

表 1 青光眼基本情况描述及单因素分析

变量	数量（例）	构成比（%）	住院总费用（元）	T/χ^2	P
年龄（岁）				8.015	<0.001
<30	32	1.5	10 014.68 ± 17 442.25		
30 ~ 45	87	4.1	6 201.35 ± 5 654.48		
46 ~ 60	369	17.3	7 064.31 ± 3 495.86		
>60	1 649	77.2	7 237.14 ± 2 945.97		
性别				29.385	<0.001
男	914	42.8	6 940.39 ± 3 782.90		
女	1 223	57.2	7 405.76 ± 3 854.07		
出院情况				16.647	<0.001
治愈	1 968	92.1	7 307.44 ± 3 724.89		
其他	169	7.9	6 033.82 ± 4 747.10		
住院天数（天）				46.911	<0.001
≤2	1 993	93.3	7 021.93 ± 3 089.71		
>2	144	6.7	9 764.20 ± 8 890.91		
是否有并发症				157.331	<0.001
无	485	22.7	5 682.67 ± 6 244.91		
有	1 652	77.3	7 654.15 ± 2 580.89		
医保类型				81.024	<0.001
城镇居民医保	103	4.8	6 953.04 ± 3 301.48		
城镇职工医保	1 795	84.0	7 176.12 ± 3 160.90		
全公费	3	0.1	7 728.60 ± 4 855.01		
全自费	231	10.8	6 930.49 ± 3 328.06		
商业医疗保险	5	0.2	35 868.34 ± 37 299.65		

3.2 青光眼患者住院费用影响因素的多元线性逐步回归分析

3.2.1 多元线性回归变量赋值 将单因素分析结果中所有具有显著性差异的自变量带入多元线性回归分析，分别赋值，见表 2。

表 2 变量赋值

变量	赋值
年龄	1 = <30 岁，2 = 30 ~ 45 岁，3 = 46 ~ 60 岁，4 = >60 岁
性别	1 = 男，2 = 女
出院情况	1 = 治愈，2 = 其他
住院天数	1 = ≤7 天，2 = >7 天
是否有并发症	1 = 有，2 = 无
医保类型	1 = 城镇居民医保，2 = 城镇职工医保，3 = 全公费，4 = 全自费，5 = 商业医疗保险

3.2.2 多元线性回归模型 将年龄、性别、出院情况、住院天数、是否有并发症、医保类型作为自变量，以住院总费用为因变量，进行青光眼患者住院费用影响因素的多元线性逐步回归分析，F = 38.662，P < 0.001，见表 3。除性别外，年龄、出院情况、住院天数、是否有并发症、医保类型对住院费用有显著影响。青光眼患者住院费用符合方差齐性（P < 0.001）。所有变量 VIF = 1.0 ~ 1.2，说明模型中各变量存在的共线性可能较小。分析标准化回归系数可知青光眼住院费用影响因素从大到小依次为：住院天数、是否有并发症、医保类型、年龄和出院情况。

表 3 青光眼患者住院费用影响因素的多元线性逐步回归分析

变量	非标准化系数		标准化回归系数 Beta	T	P	共线性统计	
	回归系数	标准误				容差	VIF
（常量）	4 271.464	760.644		5.616	<0.001	—	—
年龄	−417.817	137.686	−0.067	−3.035	0.002	0.860	1.163
性别	85.404	164.018	0.011	0.521	0.603	0.943	1.061
出院情况	−966.856	236.082	−0.085	−4.095	<0.001	0.990	1.010
住院天数	2 825.884	315.528	0.185	8.956	<0.001	0.992	1.008
是否有并发症	2 195.452	202.231	0.240	10.856	<0.001	0.865	1.156
医保类型	317.361	116.489	0.057	2.724	0.006	0.969	1.032

3.3 基于决策树模型的青光眼 DRG 优化分组分析

将住院总费用作为因变量，患者年龄、性别、出院情况、是否有并发症、医保类型作为自变量纳入决策树模型。采用 1∶1 划分训练和检验样本集。共有 2 个变量进入模型，生成 4 个节点，4 条决策路径，见图 1。

最终，纳入是否有并发症、出院情况作为模型不同分类节点，形成4 个 DRG 组。其中住院费用最高的为有并发症且出院情况为其他的青光眼患者，共 60 例，占纳入病例的 5.6%；费用最低的为无并发症青光眼患者，共 244 例，占纳入病例的 22.6%。DRG 优化分组各病例组除 DRG1 组内差异大，其他各病例组合 CV = 0.31 ~ 0.49，均小于 0.5，说明组内差异较小，分组较为合理。4 个病例组合中病种权重结果显示，有并发症且出院情况为治愈的 DRG3 病例组合，病种权重为 1.08，所消耗的医疗资源最多；无并发症的 DRG1 病例组合，病种权重为 0.76，所消耗的医疗资源最少，见表 4。

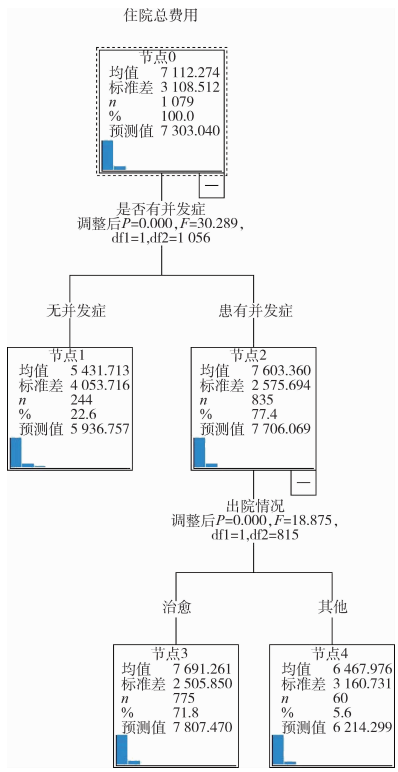

图 1 青光眼住院费用影响因素决策树模型

表 4 基于决策树模型的青光眼 DRG 优化分组结果

组别	病例组合	数量（例）	占比（%）	患者住院费用（元）		变异系数	病种权重 b
				均值	标准差		
DRG1	无并发症	244	22.6	5 431.713	4 053.716	0.75	0.76
DRG2	有并发症	835	77.4	7 603.360	2 575.694	0.34	1.07
DRG3	有并发症，出院情况为治愈	775	71.8	7 691.261	2 505.850	0.33	1.08
DRG4	有并发症，出院情况为其他	60	5.6	6 467.976	3 160.731	0.49	0.91

注：变异系数 = 住院费用标准差/均值；病种权重 b = 该组合的平均费用/所有病例的平均费用。

4 讨论

4.1 青光眼患者住院费用的主要影响因素

为加快推动国家 DRG 试点，国家医疗保障局于 2019 年 10 月印发《疾病诊断相关分组（DRGs）付费国家试点技术规范和分组方案的通知》，强调开展 DRG 试点应结合实际医疗情况，制定本地细分的 DRG 分组[13]。但实际诊疗工作中，在以诊断和治疗方式形成不同组合的基础上，需进一步考虑手术方式、病例个性化对费用的影响。医保费用的过快增长，既有疾病谱改变、人口老龄化，以及保健意识增强等客观原因，也有医疗制度设计、医院运营管理、医生利益驱动和患者盲目就诊等主观原因。为了有效控制医疗费用的过快且不合理增长，国内试行了单病种付费、按病种分值付费等医改措施。上海市第一人民医院是上海市医疗保障局 DRG 医保支付方式改革试点医疗机构，青光眼是院内重点学科眼科重点监测病种之一，年住院患者达 2 000 多人次，在上海市 23 所市级医院中病源例数排名靠前。本研究结合单因素及多元线性回归，初筛影响青光眼患者住院费用的主要因素，作为分析变量构建决策树模型，预测不同 DRG 病例住院费用标准及病种权重，推动医保支付方式改革，控制青光眼住院患者的住院费用，减少不必要的医疗资源消耗。

青光眼患者住院费用影响因素的单因素及多元线性回归分析显示，年龄、出院情况、住院天数、是否有并发症、医保类型与住院费用相关。李超等[14]对青年组、中年组、老年组 3 组病例数据进行单因素分析，提示年龄因素对住院费用的影响不容忽视。出院情况为治愈的青光眼患者住院费用高于其他出院情况患者，推断考虑治愈患者及其医生更加趋向于采取积极的治疗方案，用药、操作、检查等次数明显较多，临床疗效相比于其他患者来说相对显著，相应的住院费用也会有所上升。并发症很大程度上反映了患者病情的复杂程度，并发症越严重、数量越多，提示患者病情越复杂、治疗越困难、需要消耗越多的医疗资源，导致住院费用越高。柏杨等[15]研究发现不同医保类型患者住院费用水平差距较大，城镇职工医保患者费用高于城镇居民医保患者，最低为新型农村合作医疗患者，自费患者费用高于医保患者费用。考虑本研究中多为城镇职工医保、商业医疗保险等，新型农村合作医疗患者比例较少，不具有样本代表性。因此，提示考虑针对高龄、有并发症且出院情况为治愈的青光眼患者调整 DRG 分组测算权重。

4.2 多维度分组变量，提供青光眼 DRG 分组方案优化参考

决策树模型根据各层次分节点不同，可以有效找出影响住院费用的核心因素。本研究决策树模型预测结果显示，各 DRG 分组住院费用的 CV 为 0.33～0.75，均小于 1，说明各 DRG 分组组内同质性较好，组间差异性较大，均达到了相关规定标准。以疾病诊断为基础，需要重点考虑患者并发症、出院情况，通过特定的规则将青光眼患者分入不同组别，按照分组结果结合决策树模型预测参考标准费用及病种权重。分类节点变量将全部研究病例分为 4 个 DRG 分组方案。第一层分类节点是有无并发症，第二层分类节点是出院情况。组合 1 的住院费用和病种权重最低，说明无并发症患者所需的医疗资源支出最少。组合 3 的住院费用和病种权重最高，说明有并发症且出院情况为治愈的患者治疗难度较大，住院费用也较高。提示该病例组合的青光眼患者在住院期间可能使用较多治疗性操作和检查，一定程度上增加了医疗资源的消耗，导致住院

费用升高，这与毛英等[16]、池旋等[17]的研究结果一致。对于 DRG 付费制度，在指定青光眼住院费用时，需要考虑有无并发症和患者出院情况对医保费用的影响。本研究中，有并发症患者普遍比无并发症患者费用高约 2 000 元/例，出院情况为治愈患者普遍比其他患者费用高约 1 000 元/例。因此，对青光眼住院患者的 DRG 分组病例组合较为客观、全面地体现了患者病情严重程度和复杂程度，也进一步体现了对于医疗资源消耗的影响程度。

决策树模型提示，青光眼住院患者各 DRG 分组方案的住院费用因并发症、出院情况等多种因素影响，差异具有统计学意义。住院费用最低为 5 431.713 元，最高可达 7 691.261 元。针对不同青光眼住院患者，可以参考标准费用支付，也可以在参考值范围内根据实际情况适当增加或减少，达到控制病种住院费用、减轻患者疾病负担的目的。各组的住院费用与所有病例住院费用的比值作为各组的权重，权重越高，说明患者消耗的医疗资源越多[18]。从医疗需求来看，超过预测住院费用 5% 以内处于正常波动范围，如果费用超过 5% 的患者比例过高，提示可能出现超负荷的医疗资源，造成医疗资源浪费。因此，应关注各 DRG 分组的实际值与预测值之间的相对误差，进一步优化分组。

5 结语

本研究以 DRG 改革试点医院的青光眼患者为研究对象，运用单因素及多因素回归分析探索青光眼患者住院费用的影响因素。通过搭建决策树模型进一步对青光眼患者进行 DRG 分组方案细则的探索，以期形成合理科学的青光眼患者医疗费用分组标准，为医疗资源合理配置提供参考依据，整体 DRG 分组效果较好，为进一步推广应用 DRG 付费方式及成本管理提供了方法学依据。本研究存在一定的局限性：未进行大样本验证，仅试点单一公立医院；未考虑不同类型医疗机构的诊疗方式等。后续将进一步深入推进多家医院试点应用，进行大样本迭代验证。

作者贡献：袁筱祺负责研究设计、论文撰写；董笑负责研究设计；赵爽负责数据获取与分析。

利益声明：所有作者均声明不存在利益冲突。

参考文献

1 HUNT M A, LUND H, DELAY L, et al. DRGquant: a new modular AI - based pipeline for 3D analysis of the DRG [J]. Journal of neuroscience methods, 2022, 37 (1): 109497.
2 邓小虹. 北京 DRG 系统的研究与应用 [M]. 北京: 北京大学医学出版社, 2015.
3 LIU Y, WANG G, QIN T G, et al. Comparison of diagnosis - related groups (DRG) - based hospital payment system design and implementation strategies in different countries: the case of ischemic stroke [J]. Bioscience trends, 2024, 18 (1): 1 - 10.
4 LI Q, FAN X, JIAN W. Impact of diagnosis - related - group (DRG) payment on variation in hospitalization expenditure: evidence from China [J]. BMC Health services research, 2023, 23 (1): 688.
5 WANG H, GAO C, DANTONA C, et al. DRG - LLaMA: tuning LLaMA model to predict diagnosis - related group for hospitalized patients [J]. NPJ digital medical, 2024, 7 (1): 16.
6 马原, 黄裕翔, 李黎, 等. DRG 付费下成都市某三甲医院急性白血病患儿住院费用分析 [J]. 医学与社会, 2024, 37 (10): 85 - 90.
7 张英强, 吴欣, 李静, 等. DRG 支付制度改革对老年患者就诊费用的影响与政策建议 [J]. 中国医院, 2024, 28 (5): 55 - 58.
8 葛龙, 张南, 王月. 某三甲医院基于 DRG 恶性肿瘤患者住院费用结构研究 [J]. 中国医院, 2023, 27 (1): 49 - 52.
9 王文君, 贾晓倩, 周道平, 等. DRG 高倍率病例与正常倍率病例差异及住院费用灰色关联分析 [J]. 中国医院, 2023, 27 (5): 55 - 58.
10 蹇思慧, 陈晓雯. 肝脏手术病例的 DRG 分组效果及费用影响因素研究——以福建某三甲医院为例 [J]. 卫生软科学, 2024, 38 (1): 74 - 78.
11 李洪超. 商业银行房地产信贷风险的度量与预测——基于多元线性回归分析方法 [J]. 上海房地, 2023 (11): 40 - 45.
12 苏昂. 用于集合滤波数据同化的自适应协方差膨胀方法研究 [D]. 天津: 天津大学, 2021.
13 黄琳, 李世远. 皮肤病医院门诊抗菌药物应用情况调查分析 [J]. 中国医院用药评价与分析, 2010, 10 (7): 597 - 599.
14 李超, 盖媛媛, 潘琦, 等. 疾病诊断相关分组预定额付费制模式下年龄因素对平均住院费用的影响分析 [J]. 中华老年医学杂志, 2021, 40 (11): 1425 - 1428.
15 柏杨, 李慧, 蔡玥, 等. 2017 年中国肺癌患者医疗服务利用及住院费用性别差异分析 [J]. 中国公共卫生, 2022, 38 (9): 1209 - 1213.
16 毛英, 孙李丽, 刘芳, 等. 合并多种慢性病的缺血性心脏病患者住院费决策树分析 [J]. 中国医院统计, 2023, 30 (5): 381 - 386.
17 池旋, 涂晓贤, 连晓岚. 基于决策树模型的急性髓系白血病 DRG 分组研究 [J]. 中华医院管理杂志, 2023, 39 (2): 97 - 101.
18 林卓琛, 辛子艺, 熊莺. 运用决策树模型对某医院不稳定型心绞痛住院费用的分析 [J]. 中国病案, 2023, 24 (8): 60 - 64.