

# 基于 CHARLS 数据集的老年人群高血压患病因素研究

王少锋<sup>1\*</sup> 张凯慈<sup>1\*</sup>

(1. 广东东软学院, 广东 佛山 528225)

**摘要:** 随着中国人口老龄化程度不断加深, 老年人群中高血压的患病率持续上升。本研究基于 CHARLS 2020 年数据, 从人口学特征、生活习惯、非高血压慢性病患者情况及地区分布四个维度, 分析影响高血压的相关因素, 并探讨其与高血压之间的关联强度。研究结果显示, 多个变量与高血压存在显著关联。描述性统计表明, 年龄组别、吸烟行为以及患有心脏病、中风、糖尿病或血糖升高、血脂异常、记忆相关疾病等因素与高血压关联密切, 其中慢性疾病的占比尤为突出。通过逻辑回归模型对系数和优势比的分析进一步证实, 患有上述慢性病及吸烟者患高血压的风险明显增高。此外, 年龄作为关键因素, 其增长与高血压风险呈正向关系。然而, 由于数据存在类别不平衡问题, 可能导致部分潜在因素的影响被低估或遗漏, 影响结果的全面性与准确性。因此, 未来研究应优化数据结构, 深入挖掘更多影响因素, 以提升高血压防控策略的科学性与有效性。

**关键词:** 相关性分析; 逻辑回归; 高血压

DOI: doi.org/10.70693/cjmsr.v1i3.1740

## 1 研究背景与意义

随着我国人口老龄化加剧, 老年人高血压患病率持续上升。《中国居民营养与慢性病状况报告(2020年)》<sup>[1]</sup>显示, 18岁以上居民高血压患病率达27.5%, 其中60岁以上人群高达59.2%, 患者总数超2.45亿, 疾病负担日益加重。高血压不仅表现为血压升高, 更可导致心脑血管、肾脏、眼部等多系统损害, 是心肌梗死、卒中、心力衰竭等重大疾病的重要诱因。其发病与遗传及不健康生活方式密切相关, 高盐饮食、吸烟饮酒、缺乏运动等因素显著增加风险。然而, 我国居民对高血压的知晓率、治疗率和控制率仍较低, 防控形势严峻。

本研究基于 CHARLS 2020 年数据, 系统分析老年人群高血压的患病现状及其影响因素, 旨在揭示关键风险因子, 为制定科学、可行的预防干预策略提供实证依据, 对提升老年群体健康水平具有重要意义。

## 2 研究综述

国内外学者对老年人高血压影响因素进行了广泛研究。国外研究中, Baozhen Dai<sup>[2]</sup>发现加纳城市居民比农村居民高血压风险更高; Sri Hari T<sup>[3]</sup>指出吸烟、饮酒、家族史及糖尿病是城市老年人高血压显著风险因素; Batubara<sup>[4]</sup>证实体力活动不足、高钠/脂肪摄入及低钾摄入显著增加高血压风险。国内研究方面, 董焯华<sup>[5]</sup>分析医疗保障、家庭结构对高血压管理的影响; 刘晓迪<sup>[6]</sup>利用 CHARLS 数据发现肥胖、超重、高血脂和高血糖是主要危险因素, 而女性、在婚、规律活动为保护因素; 袁姣<sup>[7]</sup>研究显示45岁及以上人群高血压患病率达34.38%, 危险因素包括男性、高龄、城市居住等; 唐梦龄<sup>[8]</sup>发现睡眠质量差与高血压患病率增加相关; 粟丽<sup>[9]</sup>研究表明老年人高血压知晓率64.6%、控制率24.9%, 影响因素包括年龄、居住地、生活方式等。使用逻辑回归方法的研究包括袁姣<sup>[7]</sup>、刘晓迪<sup>[6]</sup>、粟丽<sup>[9]</sup>等。综合来看, 年龄、居住地、生活方式、经济状况等因素对高血压管理有重要影响, 但现有研究多聚焦单一因素, 未来需考虑多维度综合影响, 以制定更有效的公共卫生策略。

## 3 研究方法

**作者简介:** 王少锋(1988—), 男, 硕士, 广东东软学院助教, 研究方向: 医疗信息化, 慢病数据分析;

张凯慈(2003—), 女, 广东东软学院 2021 级健康服务与管理专业。

\*所标注的作者为共同第一作者

**通讯作者:** 王少锋

### 3.1 数据来源

本研究采用 2020 年 CHARLS 数据<sup>[10]</sup>, 该全国性追踪调查项目自 2011 年启动, 全面收集中国中老年人健康状况、经济安全、家庭支持及社会参与等信息, 为分析高血压影响因素提供权威数据基础。

### 3.2 分析指标定义

本文将研究的相关变量分为被解释变量与解释变量两种。

#### 3.2.1 被解释变量

基于问卷中的 DA003 项, 将人群分为两类: “高血压患者”(记为 1); 以及“非高血压患者”(记为 0)。

#### 3.2.2 解释变量

本研究的解释变量包括四类, 所有变量均采用明确的二分类或有序分类编码, 便于后续统计分析:

(a) 人口统计学变量: 年龄由 BA003 项出生年份计算得出; 性别依据 BA001 项, “1”为男性, “2”为女性; 婚姻状况依据 BA011 项, “1”表示有配偶, “0”为无配偶; 户口依据 BA009 项, “1”为农业户口, “0”为非农业户口。

(b) 生活习惯变量: 吸烟状态依据 DA046 项, “1”为吸烟, “0”为不吸烟; 饮酒状态依据 DA051 项, “1”为饮酒, “0”为不饮酒; 睡眠时间依据 DA030 项, 分为“<6 小时”(1)、“6-8 小时”(2)、“>8 小时”(3)。

(c) 慢性病患病状况: 基于 DA003 项医生诊断信息, 选取五种慢性病作为变量, 均以“1”表示患病、“0”表示未患病, 包括血脂异常、糖尿病(或血糖升高)、心脏病、中风、与记忆有关的疾病。

(d) 地区分布变量: 结合 CHARLS (2020) 中 Community ID 与 PSU 编码, 并参考 DA008 项城乡类别, 将居住地划分为农村(“0”)和城镇(“1”)。

### 3.3 研究方法

本研究以老年人高血压患病状况为因变量, 分析人口统计学特征(年龄、性别、婚姻状况、户口)、生活习惯(睡眠、吸烟、饮酒)、慢性病患病情况及城乡地区分布等多维度影响因素。采用文献研究法梳理国内外高血压流行病学特征及研究方法; 运用描述性统计法对 CHARLS 数据进行交叉表分析和卡方检验, 初步识别显著相关因素; 通过多元逻辑回归模型深入探究各因素影响程度, 重点考察回归系数、优势比(OR)、统计显著性及 ROC 曲线等指标, 科学评估各变量对老年人高血压患病风险的影响机制与作用强度。

### 3.4 分析工具

前期数据预处理部分及逻辑回归部分主要利用 Python 的 sklearn 库中的逻辑 Regression 函数以及 statsmodels 库; 描述性统计部分使用 SPSS 软件完成。

## 4 研究结果

### 4.1 描述性分析

本研究采用 CHARLS 的 2020 年调查数据, 原始数据为 19367 个样本, 剔除其中年龄小于 60 岁的受访者以及所研究的影响因素中的缺失数据, 最终纳入研究的有 1307 个样本。统计结果显示, 其中老年人高血压患者的样本为 212 人, 患病率为 16.22%。其中, 男性与女性样本数分别为 329 人和 978 人, 总体研究样本的 25.17%和 74.83%, 女性的占比较高。

#### 4.1.1 人口统计学相关变量

①不同年龄段的老年人高血压患病情况

如表 4.1, 研究将老年人分为 60-69 岁(占样本 11.2%, 患病率 13.97%)、70-79 岁(占样本 65.7%, 患病率 21.93%)和 80+岁(占样本 23%, 患病率 17.69%)三组。卡方检验显示各组高血压患病率差异显著( $\chi^2=10.648$ ,  $p<0.05$ ), 表明年龄与高血压患病存在统计学关联, 其中 70-79 岁组患病率最高。

表 4.1 我国老年人高血压患病情况的年龄段分布

组别	未患高血压(人)	患高血压(人)	$\chi^2$	p
60-69岁	739	120	10.648	0.005
70-79岁	235	66		
80岁及以上	121	26		

#### ②不同性别的老年人高血压患病情况

如表 4.2, 研究显示, 男性老年人占 25.2%(329 人), 高血压患病率 16.72%; 女性占 74.8%(978 人), 患病率 16.05%。卡方检验表明性别与高血压患病率无显著关联( $\chi^2=0.080$ ,  $p=0.777$ ), 两性间患病率差异不具统计学意义, 总体患病率为 16.22%。

表 4.2 我国老年人高血压患病情况的性别状况分布

组别	未患高血压(人)	患高血压(人)	$\chi^2$	p
男性	274	55	0.080	0.777
女性	821	157		

#### ③不同婚姻状况的老年人高血压患病情况

如表 4.3, 70.8%有配偶老年人高血压患病率 15.38%, 29.2%无配偶者患病率 17.06%。卡方检验显示婚姻状况与高血压患病率无显著关联( $\chi^2=0.279$ ,  $p>0.05$ ), 差异不具统计学意义。

表 4.3 我国老年人高血压患病情况的婚姻状况分布

组别	未患高血压(人)	患高血压(人)	$\chi^2$	p
有配偶	779	316	0.279	0.597
无配偶	147	65		

#### ④不同户口类型的老年人高血压患病情况

如表 4.4, 农业户口老年人占 73%(患病率 16.25%), 非农业户口占 27%(患病率 16.15%)。卡方检验显示户口类型与高血压患病率无显著关联( $\chi^2=0.002$ ,  $p>0.05$ ), 差异不具统计学意义。

表 4.4 我国老年人高血压患病情况的户口类型分布

组别	未患高血压(人)	患高血压(人)	$\chi^2$	p
农业	799	155	0.002	0.965
非农	296	57		

### 4.1.2 生活习惯相关变量

#### ①不同吸烟状态的老年人高血压患病情况

如表 4.5, 92.3%不吸烟老年人高血压患病率 15.17%, 7.7%吸烟者患病率 28.71%。卡方检验显示吸烟状态与高血压患病率显著相关( $\chi^2=12.570$ ,  $p<0.05$ ), 表明吸烟者高血压风险显著增高, 是重要危险因素。

表 4.5 我国老年人高血压患病情况的吸烟状态分布

组别	未患高血压(人)	患高血压(人)	$\chi^2$	p
吸烟	72	29	12.570	0.000
不吸烟	1023	183		

#### ②不同饮酒状态的老年人高血压患病情况

如表 4.6, 74.2%不饮酒老年人高血压患病率 17.22%, 25.8%饮酒者患病率 13.35%。卡方检验显示饮酒状态与高血压患病率无显著关联( $\chi^2=2.747$ ,  $p>0.05$ )。数据显示不饮酒组患病率高于饮酒组, 但差异不具统计学意义。

表 4.6 我国老年人高血压患病情况的饮酒状况分布

组别	未患高血压(人)	患高血压(人)	$\chi^2$	p
饮酒	292	45	2.747	0.097
不饮酒	803	167		

#### ③不同睡眠时间的老年人高血压患病情况

如表 4.7, 睡眠<6 小时老年人占 30.6%(患病率 18.75%), 6-8 小时占 42.8%(16.25%), >8 小时占

26.5%(13.26%)。卡方检验显示睡眠时间与高血压患病率无显著关联( $\chi^2=4.127$ ,  $p>0.05$ ), 三组间患病率差异不具统计学意义。

表 4.7 我国老年人高血压患病情况的睡眠时间分布

组别	未患高血压(人)	患高血压(人)	$\chi^2$	p
小于 6h	325	75	4.127	0.127
6-8h	469	91		
大于 8h	301	46		

#### 4.1.3 非高血压慢性病患病情况相关变量

##### ①血脂异常

如表 4.8, 血脂异常老年人占 8.7%(患病率 47.37%), 无异常者占 91.3%(患病率 13.24%)。卡方检验显示血脂状态与高血压患病率显著相关( $\chi^2=89.167$ ,  $p<0.05$ ), 表明血脂异常者高血压风险显著增高, 是高血压的重要风险因素。

表 4.8 我国血脂异常的老年人高血压患病情况分布

组别	未患高血压(人)	患高血压(人)	$\chi^2$	p
血脂异常	60	54	89.167	0.000
无血脂异常	1035	158		

##### ②糖尿病(或血糖升高)

如表 4.9, 有糖尿病/血糖升高老年人占 5.4%(高血压患病率 52.11%), 无糖尿病者占 94.6%(患病率 14.16%)。卡方检验显示糖尿病状态与高血压患病率显著相关( $\chi^2=71.174$ ,  $p<0.05$ ), 表明糖尿病患者高血压风险显著增高, 是高血压的重要风险因素。

表 4.9 我国糖尿病或血糖升高的老年人高血压患病情况分布

组别	未患高血压(人)	患高血压(人)	$\chi^2$	p
有糖尿病	34	37	71.174	0.000
无糖尿病	1061	175		

##### ③心脏病

如表 4.10, 有心脏病老年人占 7.6%(高血压患病率 52.53%), 无心脏病者占 92.4%(患病率 13.25%)。卡方检验显示心脏病与高血压患病率显著相关( $\chi^2=103.890$ ,  $p<0.05$ ), 表明心脏病患者高血压风险显著增高, 是高血压的重要风险因素。

表 4.10 我国患心脏病老年人的高血压患病情况分布

组别	未患高血压(人)	患高血压(人)	$\chi^2$	p
有心脏病	47	52	103.890	0.000
无心脏病	1048	160		

##### ④中风

如表 4.11, 中风老年人占 1.5%(高血压患病率 57.89%), 无中风者占 98.5%(患病率 15.61%)。卡方检验显示中风与高血压患病率显著相关( $\chi^2=24.641$ ,  $p<0.05$ ), 表明中风患者高血压风险显著增高, 是高血压的重要风险因素。

表 4.11 我国患中风老年人的高血压患病情况分布

组别	未患高血压(人)	患高血压(人)	$\chi^2$	p
中风	8	11	24.641	0.000
无中风	1087	201		

##### ⑤与记忆有关的疾病

如表 4.12, 有记忆相关疾病老年人占 3.6%(高血压患病率 40.43%), 无此病者占 96.4%(患病率 15.32%)。卡方检验显示记忆疾病与高血压患病率显著相关( $\chi^2=21.019$ ,  $p<0.05$ ), 表明有记忆疾病者高血压风险显著增高, 是高血压的重要风险因素。

表 4.12 我国与记忆有关的疾病老年人的高血压患病情况分布

组别	未患高血压(人)	患高血压(人)	$\chi^2$	p
有记忆疾病	28	19	21.019	0.000
无记忆疾病	1067	193		

#### 4.1.4 地区分布相关变量

如表 4.13, 农村老年人占 62.1%(高血压患病率 16.38%), 城镇老年人占 37.9%(患病率 15.96%)。卡方检验显示居住地与高血压患病率无显著关联( $\chi^2=0.040$ ,  $p>0.05$ ), 差异不具统计学意义, 但农村患病率略高于城镇。

表 4.13 我国老年人高血压患病情况城乡分布

组别	未患高血压(人)	患高血压(人)	$\chi^2$	p
农村	679	133	0.040	0.842
城镇	416	79		

## 4.2 逻辑回归

### 4.2.1 模型的构建与分析

#### ①变量准备

数据预处理中, 先合并相关列再删除缺失值以处理大量缺失数据。变量选择方面, 从 14 个慢性病变量中筛选出卡方检验  $\chi^2>20$  且  $p<0.05$  的 5 个强相关变量(血脂异常、糖尿病、心脏病、中风、记忆相关疾病)。对分类变量进行数值编码, 如性别(男=1, 女=0)、居住地(城镇=1, 农村=0)等。尽管描述性统计显示饮酒、性别、户口、居住地、婚姻状况及睡眠时间无显著性, 但参考陈娇<sup>[11]</sup>研究, 仍将其纳入模型。最终确定模型输入包括 5 种慢性病、吸烟饮酒习惯、人口统计学特征及睡眠时间, 以全面评估高血压影响因素。

#### ②多元逻辑回归分析

为了探究不同因素对老年人患高血压的影响, 采用了多元逻辑回归模型。具体步骤如下:

第一, 数据划分: 将原始数据集划分为训练集和测试集, 比例为 8:2, 以保证模型评估的客观性。

第二, 模型训练: 利用 sklearn 库中的逻辑 Regression 函数训练模型, 并使用 statsmodels 库来获取更详细的统计信息。

第三, 模型评估: 通过准确率、ROC AUC 得分、混淆矩阵以及分类报告对模型性能进行评估。

### 4.2.2 模型的结论

如表 4.14, 模型测试准确率 84.7%, ROC AUC 为 0.7075, 整体预测能力一般。由于数据不平衡, 模型对无高血压预测准确度高, 而对高血压患者识别能力有限。如表 4.16, 逻辑回归分析显示: 心脏病为最高风险因素(系数 1.57, OR=4.78), 其次为糖尿病、血脂异常、中风和吸烟(OR>1); 婚姻状况、居住地、饮酒和户口呈负相关(OR<1), 可能具保护作用; 年龄、性别和睡眠时间影响相对较小。结果表明慢性疾病与吸烟是高血压的主要危险因素, 如表 4.105, 模型对少数类识别能力不足, 可能受数据不平衡和特征局限性影响, 需进一步优化以提高对高血压患者的预测准确性。

表 4.14 逻辑回归性能表

训练集	测试集	准确率	混淆矩阵	ROC AUC
1045	262	0.847	[[214 3] [37 8]]	0.708

表 4.15 逻辑回归分类报告表

指标	精确率	召回率	F1 值	支持数
类别 0	0.85	0.99	0.91	217
类别 1	0.73	0.18	0.29	45
准确率			0.85	262
宏平均值	0.79	0.58	0.60	262
加权平均值	0.83	0.85	0.81	262

表 4.16 逻辑回归模型系数表

特征	系数	优势比 (OR)	OR 排序
血脂异常	1.27316	3.087358	3
糖尿病	1.194838	3.303024	2
心脏病	1.565320	4.784207	1
中风	1.071617	2.920097	4
记忆疾病	0.583749	1.792746	6
是否吸烟	1.033432	2.810695	5
是否饮酒	-0.439919	0.644089	12
性别	0.029272	1.029705	8
户口	-0.487764	0.613998	13
居住地	-0.182325	0.833330	11
婚姻状况	-0.031787	0.968713	10

### 4.2.3 模型的改进

针对逻辑回归模型中出现的类别不平衡问题尝试使用 SMOTE 算法解决。首先，检查类别分布。在应用 SMOTE 前，检查训练集中目标变量的类别分布，类别 0 和类别 1 分别 875 例和 170 例。其次，初始化 SMOTE 并对训练集中类别 1 的样本进行过采样，生成新的平衡数据集（仅对训练集进行过采样，测试集保持不变）。接着，训练模型。使用过采样后的训练集训练分类模型。最后，再次对模型进行评估。使用测试集评估模型性能，包括准确率、召回率、F1 分数、ROC AUC 等指标以及绘制 ROC 曲线以直观展示模型的区分能力。

根据输出结果，改进后的模型在测试集上的准确率为 74.81%，尽管比改进前下降了 9.92%，但对少数类（患有高血压）的识别能力有所提升。ROC AUC 得分为 0.7006，与改进前的 ROC AUC 得分相差较小，说明模型的区分能力总体表现区别不大。

使用 SMOTE 后，模型对于少数类（患有高血压）的识别能力增强，混淆矩阵变为[[178 39], [27 18]]。

使用 SMOTE 后，类别 1（患有高血压）的召回率从 0.18 提升到了 0.40，表明我们的过采样措施使得模型能识别更多真实高血压病患者，改进漏检问题。

在系数分析和优势比方面，系数均小于 1、OR 值均降低且优势比排名不变，其中心脏病的系数为 0.51，优势比为 1.66，婚姻状况、居住地、是否饮酒和户口等因素的系数依旧为负值，其优势比均小于 1。

## 5 讨论与建议

基于对逻辑回归模型的分析结果，吸烟是高风险因素，与 Baozhen Dai<sup>[2]</sup>和 Sri Hari T<sup>[3]</sup>的研究结论一致，他们都指出吸烟对高血压有显著影响。心脏病、糖尿病或血糖升高、血脂异常显著增加高血压风险与 Batubara<sup>[4]</sup>、刘晓迪<sup>[6]</sup>、袁姣<sup>[7]</sup>的研究指出慢性疾病是高血压的显著因素相符。

董烨华<sup>[5]</sup>提到的家庭结构和社会交往等因素对高血压管理的重要性，虽然这不是直接的风险因素，但与本研究中提到婚姻状况对高血压有一定影响的观点相符。虽然本研究中睡眠时间的影响相对较小，但也可能在一定程度上影响高血压的发生，与唐梦龄<sup>[8]</sup>得出睡眠质量差与高血压患病率增加有关的研究结论相似。

针对以上内容，可以提出以下几项对策建议来改善老年人高血压的预防和管理：

加强慢性病管理。鉴于心脏病、糖尿病、血脂异常、中风等慢性疾病显著增加了患高血压的风险，其心脏病和中风可能是由长期高血压导致，应加强对这些疾病的早期检测和有效管理。建议进行定期的健康检查，不仅可以帮助早期发现并及时治疗相关疾病，还能通过定期测量血压来监控高血压的情况。如果确实存在高血压，需要服用药物控制血压水平，以此预防因高血压引起的各种心血管疾病及其他相关健康问题，从而降低这些病症的总体发生率。

生活方式干预。鼓励老年人采用健康的生活方式，包括但不限于戒烟限酒、均衡饮食以及适量运动等。特别是吸烟被确认为是高血压的重要风险因素，因此需要特别重视烟草使用的减少或停止。

其中饮酒在本文研究中的显著性虽不强,但也是高血压的危险因素<sup>[7]</sup>,需要限制饮酒量。

社会支持系统的强化。考虑到婚姻状况等因素对高血压具有一定的保护作用,增强老年人的社会支持网络,例如通过社区活动促进老年人之间的互动,提升他们的心理健康状态,也可能有助于降低高血压的风险。

综上所述,针对老年人群体,特别是那些已知患有心脏病、糖尿病或其他心血管疾病的个体,应加强高血压的预防措施,而且控制慢性病的发展可能是关键的干预方向。

## 6 结语

本研究通过逻辑回归模型分析揭示了高血压的关键影响因素。针对数据不平衡问题,采用 SMOTE 算法进行优化,结果显示改进后模型对高血压患者的识别能力显著提升,类别 1 召回率从 0 增至 45%,有效缓解了漏检问题,但准确率下降至 54%,ROC AUC 值维持在 0.53,表明模型区分能力仍有限。研究证实慢性疾病(心脏病、糖尿病、血脂异常等)与吸烟是高血压的重要风险因素,而饮酒和户口类型可能具有保护作用,但需进一步验证。年龄作为不可忽视的因素,与高血压风险呈正向关系。本研究受限于数据不平衡及样本特征局限,可能导致部分潜在因素影响被低估。未来研究应探索更有效的数据平衡方法,扩大变量范围,尝试随机森林、神经网络等多元模型比较,并结合临床指标构建更精准的预测体系。此外,建议设计专门针对高血压影响因素的问卷,纳入更多行为与环境变量,为制定个性化干预策略提供更科学的依据,从而提升老年人群高血压防控效果。

### 参考文献:

- [1] 国家卫生健康委员会,国家疾病预防控制中心,国家癌症中心,等.中国居民营养与慢性病状况报告(2020年)[R].北京:北京大学医学出版社,2020.
- [2] DAI B, ADDAI-DANSOH S, NUTAKOR J A, et al. The prevalence of hypertension and its associated risk factors among older adults in Ghana[J]. *Frontiers in Cardiovascular Medicine*, 2022, 9: 990616.
- [3] HARI T S, SUDHA T S, VARGHESE A M, et al. A study of risk factors and complications in elderly hypertensive subjects[J]. *Journal of Family Medicine and Primary Care*, 2021, 10(6): 2230-2234.
- [4] BATUBARA F R, WANTIAN SULING F R. The Relationship of Physical Activity and Diet to the Incidence of Hypertension in the Elderly[J/OL]. *International Journal of Science and Healthcare Research*, 2022, 7(4): 79-92.
- [5] 董烨华;冯文.我国老年人高血压管理情况的社会影响因素分析[J].*中国全科医学*, 2017, 20(34): 4310-4314,4321.
- [6] 刘晓迪;修璟威;李欣阳;崔庆霞;李望晨;王在翔.中国中老年人群高血压患病影响因素分析[J].*中国公共卫生*, 2018, 34(6): 795-797.
- [7] 袁姣;武青松;雷枢;叶霖.我国中老年人群高血压流行现状及影响因素研究[J].*中国全科医学*, 2020, 23(34): 4337-4341.
- [8] 唐梦龄,魏芳,张华芳,等.老年人群睡眠与高血压的关联研究[J].*中华流行病学杂志*, 2021, 42(7): 1188-1193.
- [9] 粟丽;赵慧明;陈婷;夏浩然;王亚莉.我国老年人高血压知晓率、控制率、服药率及其影响因素分析——基于 2018CLHLS 数据分析[J].*现代医学*, 2024, 52(1): 134-142.
- [10] 赵耀辉,陈欣欣,王亚峰,孟琴琴,薄海,陈川,···周红艳.中国健康与养老追踪调查第五轮(2020)用户手册[M].北京:北京大学国家发展研究院,2023.  
[https://charls.charlsdata.com/Public/ashelf/public/uploads/document/2020-charls-wave5/application/CHARLS\\_2020\\_User\\_Guide\\_Chinese.pdf](https://charls.charlsdata.com/Public/ashelf/public/uploads/document/2020-charls-wave5/application/CHARLS_2020_User_Guide_Chinese.pdf)
- [11] 陈娇.我国老年人高血压现状及其影响因素分析[D/OL].南京邮电大学,2021.

## Influencing Factors of Hypertension in the Elderly: A CHARLS-Based Study

Shaofeng Wang<sup>1\*</sup>, Kaici Zhang<sup>1\*</sup>

<sup>1</sup> Neusoft Institute Guangdong, Foshan, Guangdong 528225, China

*\*These authors contributed equally to this work*

**Abstract:** As China's population continues to age, the prevalence of hypertension among the elderly is steadily increasing. This study, based on CHARLS 2020 data, analyzes factors influencing hypertension from four dimensions—demographic characteristics, lifestyle habits, non-hypertensive chronic conditions, and regional distribution—and explores their association strength with hypertension. Results show significant associations between multiple variables and hypertension. Descriptive statistics indicate that age group, smoking behavior, and comorbidities such as heart disease, stroke, diabetes or elevated blood glucose, dyslipidemia, and memory-related diseases are closely linked to hypertension, with chronic conditions being particularly prominent. Logistic regression analysis of coefficients and odds ratios further confirms that individuals with these chronic diseases or who smoke face significantly higher risks of hypertension. Additionally, age is a key factor, with risk increasing with advancing age. However, due to data class imbalance, the impact of some potential factors may be underestimated or overlooked, affecting the comprehensiveness and accuracy of the results. Therefore, future research should optimize data structure and explore additional influencing factors to enhance the scientific rigor and effectiveness of hypertension prevention and control strategies.

**Keywords:** Correlation analysis, Logistic regression ; Hypertension