

基于小域估计的大同农村老年人乡镇层级关节 病患率的统计分析

韩昊栋, 李华鹏, 王珂琛

山西大同大学 数学与统计学院, 山西 大同 037009

DOI:10.61369/ASDS.2025060022

摘要 : 慢性病是严重威胁我国农村基层老年人健康的一类疾病, 已成为影响国家经济社会发展的重大公共卫生问题。本文基于大同农村地区60岁以上老年人慢性病调查数据, 结合第七次人口普查数据, 对基层乡镇关节病的平均患病率进行了分析研究, 分别使用合成估计与Logistic混合效应模型, 得到了广灵县、天镇县各乡镇关节病平均患病率的估计, 通过结果比较, 绘制了广灵、天镇两县关节病疾病地图, 为大同市基层慢性病防控提供一定的实证参考。

关键词 : 农村老年人; 关节病; 小域估计; Logistic混合效应模型

A Statistical Analysis of Township Level Prevalence of Arthropathy Among Rural Elders in Datong Based on Small Area Estimation

Han HaoDong, Li Huapeng, Wang Kechen

School of Mathematics and Statistics, Shanxi Datong University, Datong, Shanxi 037009

Abstract : Chronic diseases have become a major public health problem that affects the economic and social development of the country, which threaten the health of the elders at the grassroots level in China. This paper analyzes and researches the average prevalence of joint diseases in grassroots townships based on the survey data of chronic diseases of the elders over 60 years old in the rural areas of Datong, combined with the data of the seventh population census, and obtains the estimation of the average prevalence of joint diseases in the townships of Guangling and Tianzhen counties by using the synthetic estimation and the Logistic mixed effects model, respectively, and draws the joint disease maps of the counties of Guangling and Tianzhen through the comparison of the results. The results of this paper provide some empirical references for the prevention and control of chronic diseases at the grassroots level in Datong City.

Keywords : the rural elders; arthropathy; small area estimation; Logistic mixed effects model

引言

关节病在老年人群中是常见的多发慢性病之一, 国家卫生健康委发布的数据显示, 我国骨关节炎的总患病率达15%, 山西居民慢性病与营养状况报告指出山西省居民尤其是老年人重点慢性病患病率较高^[1]。研究老年人群关节病患病现状的文献有很多, 如曹裴娅等基于中国老年健康影响因素调查数据, 研究了中国老年人群关节炎患病现状及其影响因素^[2], 胡鲲等基于黔南地区农村老年人群的抽样调查, 研究了贵州黔南地区老年人群类风湿关节炎的患病现状^[3], 黄河等依据地域差异, 将内陆和海岛老年人分类, 对比这两类人群骨关节炎患病现状的差异^[4]。上述文献均是研究老年人群关节病患病现状, 除此之外, 也有一些研究社会因素对老年人群关节病影响的文献, 如薛新东, 葛凯啸基于中国老年健康影响因素调查数据, 研究社会经济地位对老年人关节炎的影响^[5], 孙永胜等对天镇县60岁以上农村老年人的基本健康情况作了探索性分析^[6]。不过这些文献的研究区域多聚焦于全国范围或省、市层级, 针对乡镇层级的研究还很少见。

研究乡镇层级等较小区域的患病率常会碰到的困难是数据匮乏。例如, 在一个典型的抽样调查之中, 样本量通常只能确保对总体(大域)进行可靠的统计推断, 但会出现某些小域的样本量很少甚至没有的情况。小域估计是解决该类问题的有效方法, 小域估计源于抽

基金项目: 山西大同大学博士启动基金资助项目 [2018-B-25]。

作者简介:

韩昊栋, 山西大同大学数学与统计学院在读硕士研究生, 研究方向为政府统计;

李华鹏, 山西大同大学数学与统计学院, 博士, 副教授, 研究方向为半参数统计;

王珂琛, 山西大同大学数学与统计学院在读硕士研究生, 研究方向为政府统计。

样调查,可用于政府部门新政策(如资金分配、医疗资源区域规划)的制定等方面,小域估计方法可分为两类^[7]:一类是基于设计的方法,如直接估计法、合成估计法(synthetic estimation method)等;另一类是基于模型的方法,常见的有基于区域水平的模型和基于单元水平的模型^[8]。小域估计技术应用广泛,例如周巍等借助遥感数据信息对河北省蔚县小品种农作物种植面积进行建模估计^[9],于力超用小域估计计算中国小区域贫困指标^[10]。本文的主要目的是研究大同农村老年人乡镇层级的关节病平均患病率,为基层卫生健康工作提供一定的理论依据。

本文的结构如下。第一节是数据集介绍与探索性分析,第二节实证研究分别使用合成估计法与 Logistic 混合效应模型给出了各乡镇关节病患病率的估计,绘制了广灵、天镇两县关节病患病情况的疾病地图,第三节为讨论。

一、数据介绍与探索性分析

本文数据来自大同市民政局批准开展的一项对大同农村地区 60 岁以上老年人慢性病调查,调查方式是对广灵、天镇两县 60 岁以上老人进行入户调查,一共 10385 条记录,涉及性别、年龄、主要疾病诊断、家庭住址等变量,总计 10385 名调查对象。以天镇县为例,得到的统计数字如表 1 所示。

表 1 天镇县关节病患病人数 1

乡镇	患病人数	乡镇	患病人数
赵家沟	14	南高崖	13
谷前堡镇	7	南河堡	24
贾家屯	4	三十里铺	1
遼家湾	17	新平堡	50
马家皂	8	玉泉镇	1
米薪关	64	张西河乡	23

从表 1 中可以看到,各个乡镇之间患病人数差异较大,其中米薪关患病人数最多,有 64 人,而一些乡镇如贾家屯、玉泉镇、三十里铺患病人数均不足 5 人。广灵、天镇两县老年人中不同性别、年龄组关节病的患病情况如表 2 所示。

表 2 广灵、天镇两县关节病患病情况、性别、年龄组的列联表

年龄组	性别	是否患有关节病		合计
		是	否	
60 至 69 岁	男	72	1007	1079
	女	220	1703	1923
70 至 79 岁	男	180	2043	2223
	女	352	2377	2729
80 岁以上	男	87	1137	1224
	女	128	1079	1207
合计	男	339	4187	4526
	女	700	5159	5859

从表 2 中可以发现,70-79 岁年龄段关节病患病人数最多,在三个年龄段中,女性关节病患者数量均高于男性。为进一步探究性别、年龄与关节病患病情况之间的关系,采用列联表检验进行统计分析,经列联表独立性检验,表 2 中年龄段与关节病发病的 Pearson 统计量为 6.893, p 值为 0.032,说明不同年龄段关节病发病率差异显著。性别与关节病发病的 Pearson 统计量为 55.852, P 值为 7.814×10^{-14} ,说明二者存在显著相关。

二、实证分析

在以上描述性分析的基础上,本节进一步分析各乡镇平均患病率以及患病率的影响因素之间的关系。分别使用合成估计法与单元水平 Logistic 混合效应模型得到各乡镇关节病患病率的估计,绘制广灵、天镇两县的关节病疾病地图。各乡镇的关节病平均患病率估计的基本思想如下,

$$\hat{p}_i = \left(\sum_{j \in s_i} y_{ij} + \sum_{k \notin s_i} \hat{y}_{ik} \right) / N_i \quad (1)$$

其中 \hat{p}_i 是第 i 个乡镇关节病的患病率估计, y_{ij} 是第 i 个乡镇入样样本 s_i 中第 j 个调查对象是否患关节病, \hat{y}_{ik} 是第 i 个乡镇未入样样本中第 k 个调查对象是否患关节病的估计 N_i 是第 i 个乡镇 60 岁以上老年人数量。

首先是合成估计法,以广灵县为例,其计算步骤如下,首先分别计算入样样本中男女的患病比例 q_1, q_2 ,然后根据第七次人口普查数据计算得到广灵县的 60 岁以上老年人中男性占比 λ ,最后利用上述数据得到广灵县未入样样本中第 i 个乡镇患关节病总人数的估计 $\hat{Y}_i = (N_i - n_i) \times \lambda \times q_1 + (N_i - n_i) \times (1 - \lambda) \times q_2$,广灵县第 i 个乡镇关节病患病率估计 $\hat{p}_i = \frac{\sum_{i=1}^m \hat{Y}_i + Y_i}{N_i}$,其中 n_i 是第 i 个乡镇入样样本数。

基于单元水平的 Logistic 混合效应模型的表达式为

$$\text{logit}(p_{ij}) = \log \left(\frac{p_{ij}}{1 - p_{ij}} \right) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i, i = 1, \dots, m; j = 1, \dots, n_i, \quad (2)$$

其中 p_{ij} 是第 i 个乡镇第 j 个调查对象患关节病的概率, $\mathbf{x}_{ij} = (1, x_{ij1}, x_{ij2}, x_{ij3})^T$ 为相应调查对象的协变量,其中 x_{ij1} 代表性别, x_{ij2} 代表年龄, x_{ij3} 代表血脂是否异常, $\boldsymbol{a} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ 为待估系数, $v_i \sim N(0, \sigma_v^2)$ 为第 i 个乡镇的随机效应项, m 为 21 个乡镇数量, n_i 代表第 i 个乡镇的调查样本数。使用模型 (2) 拟合该数据集,得到的估计结果见表 3。

表 3 Logistic 混合效应模型的拟合结果

变量名称	系数估计值	标准误	p 值
截距项	-3.885	0.631	<0.001
性别	0.651	0.079	<0.001
年龄	-0.003	0.006	0.865
血脂异常	0.257	0.119	0.030

从表 3 中可以看到,在显著性水平 0.05 下,性别、血脂是否异常是统计显著的,这表明在控制其他变量不变的条件下,老年

女性患关节病概率是老年男性的1.917 ($e^{0.651}$)倍,老年女性更易患有关节病,在控制其他变量不变的情况下,血脂异常的老年人患有关节病的概率是血脂正常的老年人的1.293 ($e^{0.257}$)倍,血脂异常的老年人更易患有关节病。根据表3中的估计系数,结合协变量的总体均值进行修正,得到各乡镇关节病平均患病率的估计值。两种估计方法得到的各乡镇关节病患病率估计如表4所示。

表4 两种估计方法得到的各乡镇关节病平均患病率的结果

区域	合成估计	模型估计	区域	合成估计	模型估计
斗泉乡	0.222	0.253	南高崖	0.029	0.037
谷前堡	0.026	0.018	南河堡	0.041	0.066
壶泉镇	0.214	0.185	三十里	0.035	0.005
加斗乡	0.229	0.312	塑狐乡	0.197	0.152
贾家屯	0.034	0.026	新平堡	0.037	0.075
蕉山乡	0.219	0.230	宜兴乡	0.229	0.301
梁庄乡	0.206	0.189	玉泉镇	0.037	0.004
逯家湾	0.029	0.017	张西河	0.038	0.072
马家皂	0.038	0.037	赵家沟	0.034	0.053
米薪关	0.029	0.048	作疃乡	0.218	0.229
南村镇	0.215	0.213	均值	0.113	0.120

从表4中发现,约有40%乡镇合成估计结果与模型估计结果比较相近,如蕉山乡、南村镇。但其余乡镇合成估计结果与模型估计结果相差较大,如新平堡镇合成估计结果与模型估计结果相差近一倍。合成估计法虽然能够解决部分乡镇中存在的患病个体过少的问题,但是同一个县内的各乡镇患病率较为接近,如广灵县各乡镇利用合成估计法得到的患病率估计均接近20%,天镇县各乡镇利用合成估计法得到的患病率估计均接近4%。

最后对估计结果进行可视化展示,图1是广灵县、天镇县关节病疾病地图,图中关节病患病率数据来源于表4。从图1中可以发现广灵县关节病患病率要高于天镇县,同时也可以发现两县中各乡镇的关节病患病率也有着较大的区别,加斗乡与宜兴乡关节病患病率超过25%,而在天镇县内关节病患病率超过5%的乡镇只有新平堡镇、南河堡乡和张西河乡。

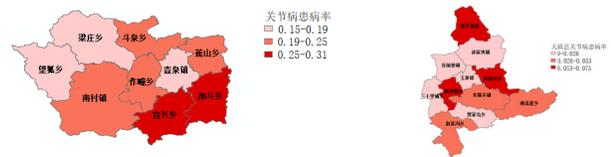


图1 广灵县关节病患病率疾病地图(左)和天镇县关节病患病率疾病地图(右)

三、讨论

本文基于小域估计技术,分别使用合成估计法与 Logistic 混合效应模型对广灵天镇两县乡镇老年人调查数据进行了统计分析,得到了各个乡镇老年人关节病患病率估计。与合成估计法相比,Logistic 混合效应模型充分利用辅助信息,模型解释性较强,解决了部分乡镇关节病患病率直接估计不可靠的问题,并给出了可靠的各乡镇关节病患病率估计。本文受数据限制,没有考虑老年人文化程度、收入水平、既往史等影响因素,我们将这留作进一步研究。

参考文献

- [1]《山西居民慢性病与营养状况报告(2020年)》首次发布 [EB/OL].[2025-7-15]. https://www.shanxi.gov.cn/ywdt/sxyw/202009/t20200902_6061315.shtml.
- [2]曹裴娅,吴侃,钱佳慧,等.中国中老年人关节炎患病现状及其影响因素分析[J].四川大学学报(医学版),2017,48(02): 268-271.
- [3]胡鲲,苏军,陈新春,等.贵州黔南地区20-79岁农村居民类风湿关节炎患病现状及危险因素[J].中国公共卫生,2019,35(07): 813-817.
- [4]黄河,田慧敏,李波,等.内陆与海岛地区老年人慢性疾病状况及影响因素分析[J].吉林大学学报(医学版),2016,42(01): 168-176.
- [5]薛新东,葛凯啸.社会经济地位对我国老年人健康状况的影响——基于中国老年健康影响因素调查的实证分析[J].人口与发展,2017,23(02): 61-69.
- [6]孙永胜,杨德兵,马俊,等.农村老年人中医体质辨识及体质影响因素分析[J].健康研究,2021,41(03): 241-245.
- [7]Pfeffermann D. New important developments in small area estimation[J]. Statistical Science,2013,28: 40-68.
- [8]Rao JNK, Molina I. Small area estimation, 2nd Ed.[M]. Hoboken, Wiley, 2015.
- [9]周巍,朱荣,张锦水.遥感辅助的农作物种植面积小域估计方法研究[J].统计研究,2015,32(07): 81-86.
- [10]于力超.小区域贫困指标估计方法的比较与模拟[J].统计与决策,2020,36(11): 32-36.