

DOI:10.13350/j.cjpb.221107

• 论著 •

青藏高原东部地区人体棘球蚴病流行影响因素的空间回归分析

施丹丹, 伍卫平*, 黄嫣, 韩帅, 谢宜

(中国疾病预防控制中心寄生虫病预防控制所(国家热带病研究中心), 国家卫生健康委员会寄生虫病原与媒介生物学重点实验室, 世界卫生组织热带病合作中心, 国家级热带病国际联合研究中心, 上海 200025)

【摘要】 目的 研究青藏高原东部地区人体棘球蚴病流行与环境、生物和社会经济因素之间的关系, 为优化棘球蚴病防控策略提供科学依据。方法 收集2020年青藏高原东部地区30个县/市棘球蚴病患者数据及环境、生物、社会经济因素资料, 计算人群棘球蚴病患病率和相关因素指标并进行空间回归分析。结果 以人群棘球蚴病患病率作为因变量, 各影响因素为自变量进行空间回归分析, 其中普通二乘回归模型法显示模型残差具有空间自相关性(Moran's $I=0.069, P<0.05$); 选择空间滞后模型进行拟合, 空间回归系数有统计学意义($\rho=0.490, P<0.05$), 拟合优度良好。模型拟合分析显示, 当地人群棘球蚴病患病率与牧草地面积占比、家畜患病率及犬棘球蚴粪抗原阳性率呈正相关, 且回归系数差异均有统计学意义(均 $P<0.05$)。结论 青藏高原东部地区人体棘球蚴病流行与环境、生物和社会经济因素等有关, 牧草地面积占比、家畜患病率、犬棘球蚴粪抗原阳性率是棘球蚴病流行的重要影响因素。采用空间回归模型分析人体棘球蚴病流行及其影响因素的效果优于传统回归模型。

【关键词】 棘球蚴病; 影响因素; 空间回归模型

【中图分类号】 R532.32

【文献标识码】 A

【文章编号】 1673-5234(2022)11-1273-05

[*Journal of Pathogen Biology*. 2022 Nov. ;17(11):1273-1277.]

Spatial regression analysis on the influencing factors of echinococcosis prevalence in eastern Qinghai-Tibet Plateau

SHI Dan-dan, WU Wei-ping, KUI Yan, HAN Shuai, XIE Yi (*National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention (Chinese Center for Tropical Diseases Research); NHC Key Laboratory of Parasite and Vector Biology; WHO Collaborating Centre for Tropical Diseases; National Center for International Research on Tropical Diseases, Shanghai 200025, China*)

【Abstract】 **Objective** To explore the relationship between the prevalence of human echinococcosis and the influencing factors in the eastern Qinghai-Tibet Plateau, so as to provide scientific evidence for improving the echinococcosis control.

Methods the data of human echinococcosis and the environmental, biological and socioeconomic factors in 30 counties in the eastern Qinghai-Tibet Plateau in 2020 were collected to obtain the prevalence of echinococcosis and to fit the spatial regression models. **Results** The prevalence of population echinococcosis was used as the dependent variable and the influencing factors were used as independent variables for spatial regression analysis. The ordinary linear regression model showed that the residuals were autocorrelated (Moran's $I=0.069, P=0.037$) and the spatial regression was necessary.

The spatial lag model was selected according to the results obtained by Lagrange multiplier statistics (LM-lag, $t=7.125, P<0.05$; Robust LM-lag, $t=9.191, P<0.05$). The spatial parameter ρ of spatial lag model was significant ($\rho=0.490, P<0.05$) and the prevalence rate of echinococcosis were positively correlated with the proportion of pasture, the prevalence rate of livestock, and the positive rate of canine Echinococcus (all $P<0.05$). **Conclusion** When explore the relationship between the factors and echinococcosis, the results derived from spatial regression models are superior to classical regression models. The major factors which affect echinococcosis are the proportion of pasture, the prevalence rate of livestock, and the positive rate of canine Echinococcus.

【Key words】 echinococcosis; influencing factors; spatial regression models

*棘球蚴病是棘球蚴的幼虫(棘球蚴)感染引起的人兽共患寄生虫病, 在全球范围内广泛分布^[1]。棘球蚴病传播和流行过程复杂, 受海拔高度、植被覆盖度等自然因素和生产方式、生活水平、卫生资源的供给等社会因素, 以及犬、牛、羊的分布和密度等生物因素的协

同作用, 给人体健康和畜牧业造成危害, 带来沉重的经

* **【通讯作者】** 伍卫平, E-mail: wuwp@nipd.chinacdc.cn

【作者简介】 施丹丹(1997-), 女, 安徽人, 硕士在读, 主要从事于流行病学与卫生统计学, E-mail: shidd1997@163.com

济负担,是当今严重的人兽共患寄生虫病^[2]。棘球蚴病流行的影响因素主要分为自然环境因素、生物因素和社会经济因素,每个因素对棘球蚴病的影响程度不同。因此,明确棘球蚴病流行的高风险因素有利于疾病的预防和控制。

随着近年来空间流行病学研究的发展,使这些空间统计学方法广泛用于描述人体疾病和健康事件的空间分布规律、分析疾病的影响因素和重点人群、探索疾病传播模式和流行趋势、制定卫生政策和合理配置卫生资源等方面^[3]。研究表明,棘球蚴病的空间分布特征具有空间聚集性^[4],但以往关于人群棘球蚴病影响因素的研究大多应用的是传统回归方法,而空间回归分析假设统计数据具有空间依赖性和空间异质性,可以利用数据的空间属性更好地揭示疾病与影响因素之间的关系。该模型不要求数据独立,可以充分分析数据的空间属性,已广泛应用于探讨疾病的时空分布和影响因素^[5]。本研究以县(市)为单位从空间维度上利用空间回归模型探索人体棘球蚴病与环境、生物和社会经济影响因素的关系,为优化人群棘球蚴病的防治策略提供参考。

材料与方法

1 资料来源

1.1 疾病数据 选取青海省玉树州、果洛州和四川省甘孜州 30 个县/市作为研究区域,收集 2020 年青藏高原东部地区棘球蚴病监测方案中的人群棘球蚴病患病的监测数据,包括人群检查总数、患病人数等指标,计算人群棘球蚴病患病率。

1.2 影响因素数据 入选区域的高程数据和植被指数数据来源于中国科学院资源环境科学数据中心。家畜患病率、犬棘球蚴虫粪抗原阳性率和犬驱虫率来源于 2020 年棘球蚴病监测数据库。户数、牛数、羊数、犬数、牧草地面积、总体土地面积、地区生产总值和农林牧渔业产值数据来源于各州 2020 年统计年鉴,用于计算户均牛数,户均羊数,户均犬数及牧草地面积占比[(牧草地面积/总体土地面积)×100%]等相关指标。

2 方法

2.1 相关指标的计算与分析 分别计算 2020 年青藏高原东部地区 30 个县/市人体棘球蚴病患病率以及高程、植被指数、牧草地面积占比、家畜患病率、犬棘球蚴虫粪抗原阳性率、户均牛数、户均羊数、户均犬数、犬驱虫率、地区生产总值和农林牧渔业产值的均值、中位数、最大值、最小值和标准差。

2.2 空间回归分析 空间回归分析是解决在空间上相邻近的地理单元间可能存在空间依赖的问题,利用数据的空间属性,建立空间矩阵权重分析数据,更好地

揭示变量之间的关系^[6-7]。空间回归分析根据参数 ρ 和 λ 可分为普通线性回归模型(Ordinary linear regression model)、空间滞后模型(Spatial lag model, SLM)和空间误差模型(Spatial error model, SEM)^[8]。

2.2.1 建模 采用普通最小二乘回归模型(Ordinary least squares, OLS)拟合判断模型拟合后的残差是否存在空间自相关性,检验统计量为 Moran's I 系数,若 Moran's I 值的 $P < 0.05$,则残差存在空间自相关性,需要进一步采用空间回归模型进行模型拟合。依据拉格朗日乘数(Lagrange Multiplier)统计量进行空间回归模型选择^[9]。若 LM-lag 有统计学意义而 LM-err 无统计学意义,则采用空间滞后模型进行空间回归分析;若 LM-err 有统计学意义而 LM-lag 无统计学意义,则采用空间误差模型进行空间回归分析;若 LM-lag 和 LM-err 有统计学意义,则比较 Robust LM-lag 和 Robust LM-err 检验统计量来选择模型。若 Robust LM-lag 有统计学意义而 Robust LM-err 无统计学意义,则采用空间滞后模型;Robust LM-err 有统计学意义而 Robust LM-lag 无统计学意义,则用空间误差模型,如二者差异均有显著性,则选 P 值较小的统计量进行模型选择^[10]。

2.2.2 模型评价 模型评价指标有决定系数(R^2 、对数似然值(Log likelihood, LIK)、赤池信息准则(Akaike information criterion, AIC)和施瓦兹准则(Schwarz Criterion, SC)等^[11]。 R^2 和 LIK 值越大,或 AIC 和 SC 值越小,表示模型拟合效果越好。

数据描述性分析在 SPSS 19.0 软件中进行,空间模型分析过程在 ArcGIS10.1 和 Geoda 1.18 软件中进行。检验水准 $\alpha = 0.05$ 。

结果

1 人群棘球蚴病患病情况

2020 年,青藏高原东部地区 30 个县/市人群棘球蚴病患病率范围在 0.000%~9.463%之间。经正态性检验,人群棘球蚴病患病率不符合正态分布($Z = 0.560, P > 0.05$),经对数转换后符合正态分布($Z = 0.947, P > 0.05$)。

2 影响因素特征

以县级为单位,2020 年青藏高原东部地区的高程为 3 131.774~4 812.364 m,植被指数为 0.138~0.558。养牛数平均值为 11.192 只/户,养羊数平均值为 3.096 只/户,家畜棘球蚴病平均患病率为 1.228%。养犬数平均值为 0.327 只/户,犬驱虫率平均值为 95.304%,犬棘球蚴虫粪抗原阳性率均值为 0.703%。牧草地面积占比为 12.407%~98.748%,平均农林牧渔业产值为 54 195.669 万元,平均地区生

产总值为 163 378.617 万元(表 1)。

表 1 2020 年青藏高原东部 30 个县/市人体棘球蚴病流行影响因素特征分析

Table 1 The characteristics of influencing factors in 30 counties in the study area in 2020

因素 Factors	均值 Mean	中位数 Median	最小值 Minimum	最大值 Maximum	标准差 Standard Deviation
高程(m)	4220.963	4278.035	3131.774	4812.364	338.990
植被指数	0.381	0.400	0.138	0.558	0.111
家畜患病率(%)	1.228	0.000	0.000	8.667	2.494
户均牛数(只/户)	11.192	11.467	0.312	30.066	6.058
户均羊数(只/户)	3.096	1.912	0.000	13.991	3.665
犬棘球绦虫粪抗原阳性率(%)	0.703	0.481	0.000	2.435	0.757
犬驱虫率(%)	95.304	100.000	78.912	100.000	6.804
户均犬数(只/户)	0.327	0.314	0.077	0.746	0.168
牧草地面积占比(%)	50.559	53.474	12.407	98.748	0.263
农林牧渔业产值(万元)	54195.669	52956.500	11140.300	100800.000	24755.097
地区生产总值(万元)	163378.617	137354.000	38157.000	1048472.000	178582.693

3 空间回归分析

经对数转换后,以人群棘球蚴病患病率作为因变量,各影响因素为自变量进行 OLS 模型拟合,结果见表 2。其中 $R^2=0.849$, $LIK=-26.273$, $AIC=76.546$, $SC=93.361$ 。模型拟合后残差的 Moran's $I=0.069$, $P<0.05$ 。说明 OLS 模型拟合残差存在空间自相关性,不独立,随后以空间自相关为前提的空间回归模型对数据进行拟合。

表 2 人体棘球蚴病患病率与影响因素的 OLS 分析
Table 2 The OLS results of the Prevalence rate of human echinococcosis and influencing factors

变量 Variable	β	标准差 Standard Deviation	t	P
常数项	-5.48112	4.19035	-1.30803	0.20732
高程	0.000172623	0.000850776	0.202901	0.84149
NDVI	-2.74424	3.0808	-0.890753	0.38481
牧草地面积占比	2.11489	0.80239	2.63574	0.01679
家畜患病率	23.3447	7.36281	3.17063	0.00529
犬棘球绦虫粪抗原阳性率	54.4369	22.8696	2.38032	0.02856
户均犬数	0.566311	1.09003	0.519538	0.60972
农林牧渔业产值	1.44085e-005	9.70399e-006	1.4848	0.15490
GDP	-1.29378e-006	1.01824e-006	-1.2706	0.22005
犬驱虫率	-3.05578	3.3978	-0.89934	0.38035
户均牛数	0.0613739	0.0354849	1.72958	0.10082
户均羊数	-0.017716	0.0499738	-0.354506	0.72708

空间依赖性检验结果可以看出,LM-lag 统计量和 Robust LM-lag 统计量分别为 7.125、9.191,具有统计学意义(均 $P<0.05$);而 LM-error 统计量和 Robust LM-error 统计量分别为 0.251、2.317,均无统计学意

义(均 $P>0.05$)(表 3)。因此根据空间回归模型选择标准,选择空间滞后模型进行下一步的模型拟合。

表 3 人群棘球蚴病患病率与影响因素空间依赖性检验结果
Table 3 The spatial dependence test results of prevalence rate of human echinococcosis and influencing factors

Lagrange 乘数 Lagrange Multiplier	MI /DF	t 值 t -value	P 值 p -value
Lagrange Multiplier(lag)	1	7.125	0.008
Robust LM(lag)	1	9.191	0.002
Lagrange Multiplier(error)	1	0.251	0.616
Robust LM (error)	1	2.317	0.128

空间滞后模型的拟合效果为 $R^2=0.900$, $LIK=-21.099$, $AIC=68.199$, $SC=86.414$ 。根据拟合效果指标评判标准,空间滞后模型的 R^2 比 OLS 模型的 R^2 大,说明自变量更能描述因变量的波动情况。LIK、AIC、SC 值均比 OLS 模型的值小,说明拟合效果与 OLS 模型有改进。空间滞后参数 ρ 有统计学意义($P<0.05$),说明空间滞后模型更好地拟合了数据的空间滞后属性,也表明人群棘球蚴病患病率具有空间自相关性。

模型拟合分析显示牧草地面积占比、家畜患病率、犬棘球绦虫粪抗原阳性率与人群棘球蚴病患病率呈正相关,且回归系数均有统计学意义($P<0.05$)。(表 4)。

表 4 人群棘球蚴病患病率与影响因素空间误差模型分析
Table 4 The spatial error model results of prevalence rate of human echinococcosis and influencing factors

变量 Variable	β	标准差 Standard Deviation	t 值 t -value	P 值 P -value
ρ	0.489895	0.120665	4.05994	0.00005
常数项	-4.24233	2.69862	-1.57204	0.11594
高程	0.000221314	0.000536504	0.412512	0.67996
NDVI	-0.748618	1.98085	-0.377928	0.70548
牧草地面积占比	1.2929	0.537815	2.40399	0.01622
家畜患病率	19.0677	4.72818	4.03277	0.00006
犬棘球绦虫粪抗原阳性率	46.5061	14.4726	3.2134	0.00131
户均犬数	0.901139	0.687524	1.3107	0.18996
农林牧渔业产值	1.17315e-005	6.16665e-006	1.90241	0.05712
GDP	-3.89758e-007	6.85548e-007	-0.568535	0.56967
犬驱虫率	-1.97057	2.1499	-0.916588	0.35936
户均牛数	0.0412991	0.0224539	1.83928	0.06587
户均羊数	-0.00144621	0.0315292	-0.0458691	0.96341

讨论

自 20 世纪 90 年代起,空间回归分析方法被广泛应用于卫生健康相关问题的研究,尤其是传染病影响因素的研究^[7]。经典回归分析方法揭示因变量与自变量在数字上的联系,数据需要具有独立性,但流行病学数据常具有空间聚集性,因此不能更好地揭示数据本质。根据传统回归模型,空间回归模型引入空间效应

变量,充分利用数据的空间属性,并且不要求数据独立,因此具有更好的数据分析效果^[12]。陈艳艳等^[13]对湖北省血吸虫中间宿主钉螺分布的影响因素进行分析,运用空间回归模型研究钉螺密度与气象因素的关系。结果显示,钉螺的空间分布存在空间自相关性,空间滞后模型的拟合效果要优于普通线性回归模型。潘翔等^[14]对东洞庭湖君山区人群血吸虫的影响因素进行研究,使用 OLS 模型和空间回归模型分别拟合 2012 年和 2014 年人群血吸虫感染与家畜患病率、易感环境面积、易感环境长度之间的关系。结果显示,2012 年空间误差模型拟合效果更好,家畜患病率和易感环境面积与人群血吸虫患病率正相关;2014 年空间滞后模型拟合效果更好,人群血吸虫患病率与易感环境面积正相关。因此认为在进行疾病统计分析时,空间回归模型更能充分兼顾到疾病的空间依赖性和异质性,更能揭示疾病与影响因素之间的关系。

本研究利用空间回归模型中的 OLS 模型和空间滞后模型对人群棘球蚴病患病率及各影响因素进行数据模型拟合。OLS 模型结果表明变量存在空间自相关,残差不独立,因此引入了以空间自相关为前提的空间回归模型,更好地进行空间数据的分析^[15]。结果表明空间滞后模型的拟合效果指标 R^2 比 OLS 模型的 R^2 大,LIK、AIC、SC 的值均比 OLS 模型的值小,说明拟合效果与 OLS 模型有改进,表明在引入空间滞后变量后模型更好地揭示了 2020 年青藏高原东部地区人群棘球蚴病流行与其影响因素的关系。

空间滞后模型分析显示牧草地面积占比、家畜患病率、犬棘球蚴虫粪抗原阳性率与人群患病率呈正相关,且回归系数均具有统计学意义(均 $P < 0.05$)。我国棘球蚴病流行区植被以草原、灌木草丛为主,适合牛、羊等家养动物以及犬、狼、狐等食肉动物的生长繁殖,这些动物正是棘球蚴病传播主要的中间宿主和终末宿主。Craig 等^[16]在甘肃南部进行了棘球蚴病的影响因素调查,结果表明灌丛/草地占总面积 50% 以上与棘球蚴病感染存在相关性。刀吉等^[17]开展的四川省棘球蚴病流行病学研究和张静宵等^[7]进行的青海省棘球蚴病流行调查均表明牧区棘球蚴病危害大于半农半牧区大于农区。牧草地占比越大,表明该地以畜牧业占比更大,导致人群患棘球蚴病的几率变大,因此牧区依旧是棘球蚴病防控的重点。

本研究结果显示,青藏高原东部地区家畜患病率是人患棘球蚴病的重要影响因素之一。棘球蚴病的传播链主要是家畜和家犬的家养循环以及啮齿目和兔形目动物与野生犬科动物的野外循环。青藏高原东部地区的主要生产方式是畜牧业,以饲养牦牛和绵羊为主,部分地区可能缺乏集中屠宰场,存在散在屠宰的现象,

犬可能因误食家畜的病变脏器造成棘球蚴病的传播流行。青藏高原地区家畜的放牧时间长达十个月,大多处于野外环境中,而野外棘球蚴病的防控措施尚未明确,仍在探索阶段,因此家畜棘球蚴病对人体棘球蚴感染影响较为严重。何伟等^[18]在四川省石渠县棘球蚴病传播及其影响因素的分析中运用单因素线性回归分析发现,牛棘球蚴病患病率是人感染棘球蚴病的影响因素,回归系数为 0.234。青藏高原东部地区是畜牧业的主要生产基地,牧民养牛养羊数量庞大,进一步加强对家畜的免疫措施,杜绝家庭私自屠宰现象的发生,集中屠宰家畜,对病变脏器作无害化处理,是预防控制家畜棘球蚴病的关键。

本研究结果还显示,犬棘球蚴虫粪抗原阳性率也是影响人群患棘球病的重要因素。棘球蚴病的传染源主要为终末宿主犬科动物。青藏高原地区以畜牧业为主,家养犬普遍,人与犬的接触密切。Zeng 等^[19]收集了 2012 年我国非青藏高原地区的患病率数据、经济数据和畜牧业生产数据,采用负二项回归分析法分析相关因素,结果显示户均犬数与人群患病率呈正相关;单因素分析显示犬粪抗原阳性率与人群患病率相关。Yuan 等^[20]在我国西部 5 省开展了棘球蚴病横断面研究,多因素分析显示羊免疫、放养犬是保护因素,流浪犬和野生动物是危险因素。棘球蚴病的主要终宿主为犬,人和其他中间宿主可因误食环境中病犬排出的虫卵而感染棘球蚴病,因此应加强对家养犬的驱虫与管理。

空间回归模型以空间权重矩阵的形式将数据的空间关系纳入模型,更能充分地解释数据的空间自相关性,适用于具有空间关系的变量,较传统线性回归模型能更好地展示疾病与影响因素的关系。从本研究结果看,牧草地面积占比、家畜棘球蚴病患病率和犬棘球蚴虫阳性率是人群患棘球蚴病的重要影响因素,因此关于人棘球蚴病防治的重点之一是对牧区家畜棘球蚴病和犬棘球蚴虫感染的防控。加强家畜屠宰管理和免疫,逐步实行牲畜定点屠宰,加强家畜免疫,每年对当年新生存栏家畜进行疫苗接种。加大传染源控制力度,完善家犬登记管理,按户建立家犬驱虫登记卡,根据当地实际情况定期开展犬驱虫,广泛动员居民参与和配合犬驱虫工作,定期投喂犬驱虫药,并在无主犬聚集的场所或经常出没的区域投放驱虫药饵,以达到对当地人体棘球蚴病流行的有效控制。

【参考文献】

- [1] Deplazes P, Rinaldi L, Alvarez Rojas CA, et al. Global distribution of alveolar and cystic echinococcosis[J]. Adv Parasitol, 2017, 95: 315-493.
- [2] 房琦, 伍卫平. 棘球蚴病流行预测中的影响因素研究[J]. 中国病

