

研究报告

时空扫描统计量在云南省野生蕈中毒聚集性探测中的应用研究

王晓雯,赵世文,刘志涛,余思洋,陆林

(云南省疾病预防控制中心营养与食品卫生所,云南昆明 650022)

摘要:目的 探讨时空扫描统计量在野生蕈中毒聚集性探测及早期预警中的应用价值。方法 采用回顾性和前瞻性时空扫描统计量建立模型,以区/县作为基本地理单位,对云南省2011—2012年野生蕈中毒数据进行分析,结果分别与三间分布描述结果及实际监测结果进行比对。结果 基于回顾性时空聚集性分析:2011年,野生蕈中毒聚集性最强的为7月12日至7月18日文山州丘北县,理论发病人数0.051人, $LLR = 149.330, P = 0.001$;2012年,野生蕈中毒聚集性最强的为6月5日至6月18日文山州麻栗坡县和西畴县,理论发病人数0.057人, $LLR = 91.767, P = 0.001$ 。基于前瞻性时空聚集性分析,对2011年6月1日至2011年6月31日的预警结果分析,2011年6月1日至2011年7月1日均发出预警信号,昭通市盐津县在7 d内实际发病数为6例,预期发病例数为0.039例, $LLR = 24.234, P = 0.001$ 。结论 回顾性时空聚集分析可以充分利用时间和空间信息,使得探测结果更为精确,能准确有效的探测出野生蕈中毒的发病模式;前瞻性时空聚集性分析可以较为准确的进行前瞻性预警分析,在实际工作中具有重要的应用价值。

关键词:野生蕈中毒;聚集性探测;时空扫描统计量;早期预警;食物中毒;云南;食品安全

中图分类号:R155 文献标志码:A 文章编号:1004-8456(2016)03-0295-06

DOI:10.13590/j.cjfh.2016.03.004

Space-time scan statistical analysis of incident detection exploring clusters for wild mushroom poisoning in Yunnan Province

WANG Xiao-wen, ZHAO Shi-wen, LIU Zhi-tao, YU Si-yang, LU Lin

(Division of Nutrition and Food Hygiene, Yunnan Center for Disease Control and Prevention, Yunnan Kunming 650022, China)

Abstract: Objective To explore the valuable of space-time scan statistics for detecting the clustering of wild mushroom poisoning and to provide early warning for officials. **Methods** The approach is based on both retrospective and prospective scan statistics which were respectively tested with data in 2011 and 2012; and the analysis was implemented at the county level. Compare the results of traditional epidemiology analyses and retrospective analysis and compare the results of prospective analysis and real-time monitoring. **Results** Based on the retrospective space-time analysis, in 2011, the clustering time of wild mushroom poisoning was from June to August, the most significant cluster was from 12th, July to 18th, July and occurred in Qiubei County, in the Wenshan Municipal area. The number of expected was 0.051. LLR (Log Likelihood Ratio) value was 149.330 and $P = 0.001$. In 2012, the most significant cluster was from 5th, July to 18th, July and occurred in Malipo County and Xichou County. The number expected was 0.057. LLR (Log Likelihood Ratio) value was 91.767 and $P = 0.001$. When analyzing the early warning results from 1st, June to 1st, July, we found that the warning signals were set out daily from 1st, June to 1st, July. Within the 7 days scan parameter, the number of expected cases was 0.039, as calculated by the model. LLR (Log Likelihood Ratio) value was 24.234 and $P = 0.001$. **Conclusion** The retrospective space-time scan statistics could take advantage of temporal and spatial information to make detecting results more accurate and effective and prospective space-time scan statistic could be applied to provide an early warning analysis.

Key words: Wild mushroom poisoning; cluster detection; space-time scan statistics; early warning; food poisoning; Yunnan; food safety

收稿日期:2015-11-03

基金项目:云南省应用基础研究计划项目青年项目

作者简介:王晓雯 女 医师 研究方向为卫生统计学在食源性疾病预防中的应用 E-mail:wangxiaowen84@sina.com

通信作者:陆林 男 主任医师 研究方向为艾滋病流行病学和卫生管理与政策 E-mail:lulin@yncdc.cn

云南省为全国野生蕈消费量最大的省份。2011—2012年,云南省每年野生蕈中毒的发病率均为全国最高(2011年为1.39/10万人年;2012年为1.78/10万人年)。同时,在云南省近两年食物中毒的发病人群中,野生蕈中毒导致的病死率也为最高

(2011年为3.74%;2012年为6.10%)。

扫描统计量是探测局部时间和/或空间事件发生数的增加,并检验这种增加是否由于随机因素造成。扫描统计量包括回顾性扫描统计分析和前瞻性扫描统计分析,其中,回顾性扫描统计分析分为单纯空间扫描统计、单纯时间扫描统计及时空扫描统计;前瞻性扫描统计分为单纯时间扫描统计和时空扫描统计^[1-2]。回顾性分析可以探索发病模式及病因,前瞻性分析可实现实时监测和早期预警。

目前国内对野生蕈中毒数据的分析主要集中在描述性分析或简单的数据挖掘,国内文献报道中,最常见的分析方法是进行流行病学三间分布描述^[3-4],但通过该方法只能粗略了解中毒的发病概况,无法深入挖掘数据中的信息和进行聚集性的准确探测。在国外的研究中,有学者对炭疽等传染病及出生缺陷等健康问题进行了时空聚集性分析^[5-6],若将野生蕈中毒数据中的时间和空间信息结合起来,采用时空聚集性分析模型,可对发病的聚集性同时进行空间和时间定位,有利于疾病预防控制部门及时采取应对措施。本研究拟通过比较传统三间分布描述与回顾性时空聚集性分析,及前瞻性时空聚集性分析与实时监测结果,探讨时空扫描统计量能否有效应用于野生蕈中毒时空聚集性的探测、发病模式的研究及早期预警系统的建立。

1 材料与方法

1.1 数据来源

野生蕈中毒数据来源于2011—2012年食源性疾病(食物中毒)报告网络。每条记录包括家庭住址、现住址编码、诊断分类、原因食品及分类、报告日期等信息。本研究选择该系统的野生蕈中毒数据进行分析。人口数据来源于云南省2011年统计年鉴^[7]。云南省地理信息系统数据,包括省界、州市、县级以上界线等来自国家基础地理信息系统网站,地理坐标等信息来源于云南省地理信息查询系统。

1.2 方法

1.2.1 扫描统计量基本原理^[2]

本研究采用时空扫描统计量进行时空聚集性探测。时空扫描统计量的扫描窗口定义为圆柱形,圆柱形的底对应一定的地理区域,圆柱形的高对应时间。圆柱形窗口在时间和空间移动,对每一个位点,圆形地面的半径都从0连续地增加到事先规定的上限值,同时,时间长度也发生变化。最后得到大量互相重叠但形状不同的圆柱,覆盖整个探测区域,每一个圆柱是一个可能的簇(cluster)。先进行区/县级回顾性时空聚集性分析,再进行区/县级前

瞻性时空聚集性分析,后者参数选取参考回顾性时空聚集性分析结果。

1.2.2 分析基本单位选取

目前能够获取的地理信息数据和人口数据的最小行政单位为区县级,本研究以区县为基本分析单位。

1.3 模型建立软件及统计学分析

本研究使用Satscan 9.0软件^[8]完成时空聚集性探测模型的建立,采用Mapinfo实现探测结果的可视化。

2 结果

2.1 三间分布分析

2.1.1 发病概况

2011和2012年云南省129个区县野生蕈中毒报告病例共计1457人,两年报告病例数分别为640人和817人。两年的年发病率分别为1.39(1/10万人年)和1.78(1/10万人年)。2011年平均发病率大于5(1/10万人年)的地区共有5个,分别为南华县11.86(1/10万人年)、姚安县10.61(1/10万人年)、文山县11.00(1/10万人年)、丘北县15.27(1/10万人年)和蒙自县7.89(1/10万人年)。2012年平均发病率大于5(1/10万人年)的地区共有8个,依此为弥勒县11.48(1/10万人年)、新平县7.69(1/10万人年)、临翔区8.02(1/10万人年)、双江县7.34(1/10万人年)、姚安县5.05(1/10万人年)、大姚县5.11(1/10万人年)、麻栗坡县5.40(1/10万人年)和瑞丽市9.94(1/10万人年)。

2.1.2 时间分布

将2011和2012年云南省野生蕈中毒的发病情况按月份绘制线图,如图1所示。由图可见,2011和2012年全年均有发病,各年发病高峰相同,均为6~8月。这两年6~8月的发病数分别占各年全年发病总数的80.00%和85.82%。各年冬季发病人数较少或不发病。

2.1.3 空间分布

2011年发病率大于5(1/10万人年)为中部的南华县和姚安县,东部的文山县、丘北县和蒙自县。2012年发病率大于5(1/10万人年)的为中部的大姚县、姚安县和新平县,西部的临翔区、双江县和瑞丽市,东部的弥勒县和麻栗坡县,见图2、3。

2.2 时空聚集性分析

2.2.1 回顾性时空聚集性分析

首先采用回顾性空间扫描统计量对云南省129个区县2011—2012年野生蕈中毒发病数据进行空间聚集性分析。在探索性分析中,将最大

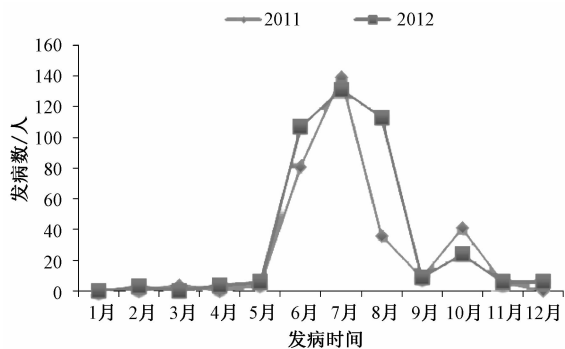


图1 2011和2012年云南省野生蕈中毒发病时间分布
Figure 1 Yunnan Province wild mushroom poisoning disease time distribution in 2011 and 2012

疾病监测网络的报告要求中,报告的时限为食物中毒结案后的7 d内上报,本研究对参数的设置也进行一种探索性的研究,因此最大 cluster 的时间长度取7 d,而 cluster 扫描半径在70 km以下依次选取,不同扫描半径进行回顾性时空聚集性分析的扫描结果见表1。

表1 回顾性时空聚集性探测各扫描半径运行时间

Table 1 Retrospective space-time aggregation to detect

the scanning radius running time		
扫描半径/km	cluster 数/个	cluster 最多覆盖的地理区域数/个
70	27	7
60	26	7
50	34	6
40	38	2
30	42	1
20	42	1
10	42	1

不同的扫描半径程序的运行时间相同。根据文献研究显示^[9-10], cluster 个数一般不宜超过地理区域总数的3/4, cluster 最多覆盖地理区域数不宜超过总数的1/10。本研究的地理区域总数为129,以上扫描半径的分析结果均符合此要求,选择居中的扫描半径,故回顾性时空聚集性分析中扫描半径确定为40 km。而前瞻性时空聚集性分析中扫描半径稍大于回顾性分析,因此下一步前瞻性时空聚集性分析的扫描半径确定为50 km。

经过回顾性时空聚集性分析,2011年,野生蕈中毒发生出现聚集性的时间为6~8月,发生地区主要为滇中、滇东和滇东北的10个区/县,聚集性最强的为2011年7月12~18日发生在文山州丘北县,实际发病人数28人,理论发病人数0.051人,对数似然比 $LLR = 149.330, P = 0.001$ 。2012年,野生蕈中毒发生出现聚集性的时间扩大到6~11月,发生地区波及除滇东北以外的大部分地区,其中,聚集性最强的为2012年6月5~18日,发生在文山州麻栗坡县和西畴县,实际发病人数共19人,理论发病人数共0.057人, $LLR = 91.767, P = 0.001$,见图4。

2.2.2 模拟实时监测系统,进行前瞻性时空聚集性分析

在野生蕈中毒的聚集性探测和早期预警中,人们关注的是当新病例不断出现的时候,能够及时发现病例异常增加而报警,探测出正在发生的聚集性。而回顾性时空聚集性探测方法是在特定的时间和空间里,判断疾病是否流行或具有暴发的趋势。因此,采用前瞻性时空扫描统计量,逐日纳入数据,模拟实时监测系统进行分析。前瞻性时空聚集性分析的参数设置为:最大 cluster 的时间长度为

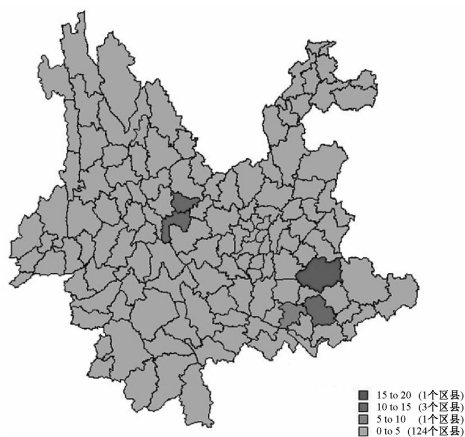


图2 2011年云南省野生蕈中毒地区分布(1/10万人年)
Figure 2 Regional distribution of Yunnan Province wild mushroom poisoning in 2011

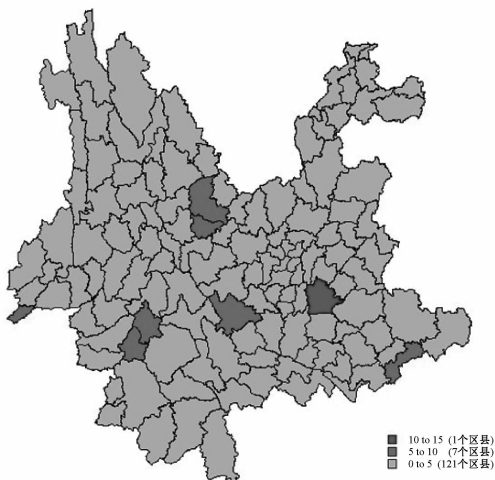


图3 2012年云南省野生蕈中毒地区分布(1/10万人年)
Figure 3 Regional distribution of Yunnan Province wild mushroom poisoning in 2012

cluster 的空间大小设定为小于总人数的50%。分析结果显示,最大 cluster 半径为73.41 km,因此,设置聚集性探测的扫描半径的上限为70 km。然后采用回顾性时空扫描统计量进行时空聚集性分析, cluster 扫描半径的上限为70 km,考虑到在食源性

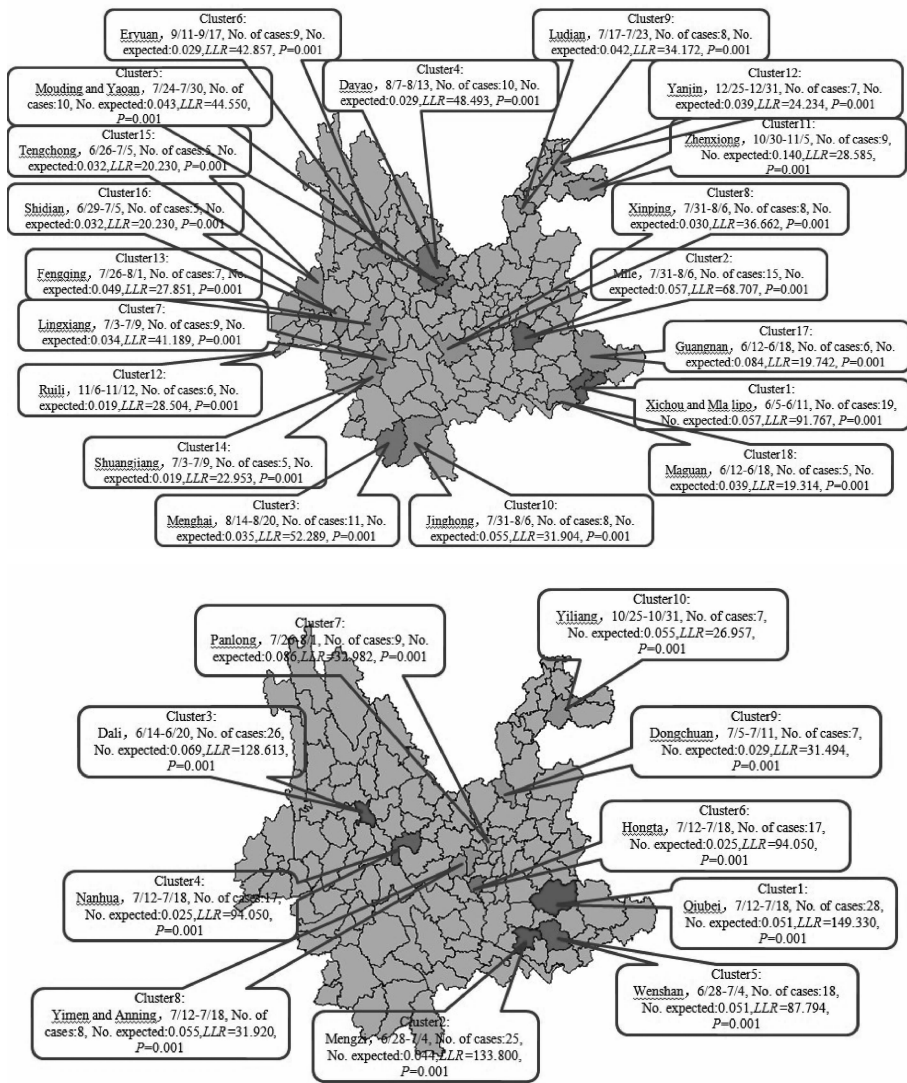


图4 2011—2012年云南省129个区/县野生蕈中毒发病回顾性时空聚集性探测结果

Figure 4 2011-2012 129 District/County in Yunnan Province wild mushroom poisoning disease retrospectively space-time clustering detection results

7 d,最大 cluster 的空间大小为扫描半径 50 km。模拟实时监测系统,对 129 个区县 2011 年 1 月 1 日—2012 年 12 月 31 日野生蕈中毒数据进行逐日前瞻性分析。即假定分析当天为 2011 年 6 月 1 日,最大 cluster 的时间长度为 7 d,则提取 2011 年 6 月 1 日、2011 年 5 月 31 日—2011 年 6 月 1 日、2011 年 5 月 30 日—2011 年 6 月 1 日、2011 年 5 月 29 日—2011 年 6 月 1 日、2011 年 5 月 28 日—2011 年 6 月 1 日、2011 年 5 月 27 日—2011 年 6 月 1 日和 2011 年 5 月 26 日—2011 年 6 月 1 日共 7 种时间长度窗口进行分析。以此类推,本研究对其中重点时段的预警结果进行进一步追踪分析。对云南省 2011 年 6 月 1 日—2011 年 6 月 31 日的预警结果进行分析,2011 年 6 月 1 日—2011 年 7 月 1 日均发出预警信号,昭通市盐津县在 7 d 内实际发病 6 人,而依据模型计算的预期发病例数为 0.039 人,LLR = 24.234,

P = 0.001,昭通市盐津县在 2011 年 6 月 1 日—2011 年 6 月 9 日可能出现野生蕈中毒的聚集性,见表 3。

3 讨论

3.1 三间分布描述结果与回顾性时空聚集性分析结果比较

三间分布描述结果发现每年的 6~8 月为野生蕈中毒发生出现聚集性的时间,云南省的中部、东部和西部为野生蕈中毒发生出现聚集性的地区。回顾性时空聚集性分析发现 2011 年野生蕈中毒发生出现聚集性的时间为 6~8 月,2012 年为 6~11 月,2011 年发生地区主要为滇中、滇东和滇东北的 10 个区/县,2012 年,发生地区波及除滇东北以外的大部分地区。此外,还发现 2011 年聚集性最强的时间为 7 月 12~18 日发生在文山州丘北县,

表3 前瞻性时空聚集性探测结果
Table 3 Prospective space-time clustering detection results

发出预警日期	cluster 序号	信号包括的天数/d	县/区	地市	实际发病数/人	预期发病数/人	LLR	P
1-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
2-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
3-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
4-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
5-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
6-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
7-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
8-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
9-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
10-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
11-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
12-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
13-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
14-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
15-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
16-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
17-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
18-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
19-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
20-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
21-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
22-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
23-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
24-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
25-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
26-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
27-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
28-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
29-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
30-Jun-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001
1-Jul-11	1	7	盐津县	昭通市	7	0.039	24.234	0.001

注:1-Jun-11 表示 2011 年 6 月 1 日,以此类推;1-Jul-11 表示 2011 年 7 月 1 日

2012 年聚集性最强的时间为 6 月 5 ~ 18 日,发生在文山州麻栗坡县和西畴县。比较两种分析方法的结果,可以看出,所探测出的聚集性发生的时间和地区分布基本一致。但回顾性时空聚集性分析可以精确定位出聚集性最强的时间和地区,比三间分布描述中的时间和地区分布更为精确。回顾性时空聚集性分析对于每一次探测到的聚集性均进行蒙特卡洛模拟运算,计算出相应的 P 值,从而排除了聚集性发生的偶然性,使得聚集性具有统计学的意义。因此,回顾性时空聚集性分析可以充分利用时间和空间信息,使得探测结果更为精确,能准确有效的探测出野生蕈中毒的发病模式。

3.2 前瞻性时空聚集性分析结果与实际监测结果的比较

本研究模拟实时监测系统,对 129 个区县 2011 年 1 月 1 日—2012 年 12 月 31 日野生蕈中毒数据进行逐日前瞻性分析。在对云南省 2011 年 6 月 1 日—2011 年 6 月 31 日进行前瞻性分析时发现

2011 年 6 月 1 日—2011 年 7 月 1 日均发出预警信号,昭通市盐津县为最有可能出现聚集性的区域,出现聚集性的时间为 2012 年 12 月 25 ~ 31 日。结合实际监测数据可以看出,2012 年 12 月 27 日,昭通市盐津县发生 2 例野生蕈中毒;2012 年 12 月 28 日,昭通市盐津县发生 5 例野生蕈中毒,5 例均死亡。发生中毒的地区均为盐津县落雁乡。前瞻性时空聚集性探测结果所发现的聚集性最强区域与实际监测结果一致。因此,前瞻性时空聚集性分析可应用于野生蕈中毒的前瞻性预警分析。

3.3 关于模型应用中的几个问题

3.3.1 模型验证与参数设置

统计学上探测到的聚集性与具有流行病学意义的爆发在概念上有所不同^[11]。聚集性探测模型的无效假设为野生蕈中毒的发生无聚集性,即病例分布为随机分布。拒绝无效假设意味着病例分布不是随机的,即存在聚集性,它可能对应于流行病学意义上的“爆发”,但也可能是散在病例正好在同

一时间 and 邻近地点发生,或可能是监测系统的信息不真实。野生蕈中毒的爆发与散发是相对的。散发指独立发生的病例,爆发指病例之间具有流行病学上的联系,因此,评价探测到的聚集性是否有流行病学意义上的爆发,需通过现场调查确定病例之间的流行病学联系。在实际工作中,需要结合现场调查及实验室检测结果,对预警结果进行评价。

扫描统计量探测聚集性的参数设置十分重要。参考国外学者在对志贺菌和大肠埃希杆菌感染性腹泻及急性疾病事件进行前瞻性时空聚集性分析时的参数设置方法^[12],在模拟实时监测时,每次分析纳入发病时间当天及当天之前的29 d的病例。这一时间长度可避免纳入太多与当前无关的信息,从而可探测到近期疫情的异常状况。在理论上,模型参数的设置仍需参考多年的野生蕈中毒监测数据以及由此反映出来的流行规律。就实际应用而言,预警结果的准确性需结合现场调查及实验室检测结果加以验证,并根据验证结果调整模型参数。

3.3.2 关于实际应用中可能遇到的问题

数据报告质量和空间信息问题。从食源性疾病(食物中毒)报告网络报告情况分析可以看出,报告不及时、数据不完整或不准确的问题仍然存在,而对食源性疾病聚集性探测和早期预警依靠于报告数据的及时性和准确性。在时空聚集性探测中,空间信息关系到病例的空间定位。本研究主要探测区县级的聚集性,以区县级作为分析单位。但在实际工作中,食物中毒的爆发只波及到街道乡镇的一个较小区域,若该区域整体食物中毒的发病水平较高时,则很难探测到爆发具体点的存在。

3.4 需要进一步研究的问题

本研究采用圆柱形扫描窗口进行聚集性探测,其地理区域形状主要是圆形或近似圆形,因此具有一定的局限性。如果实际发生的野生蕈中毒的聚集性形态为狭长的条状,此时的圆柱形扫描窗口很可能难以探测到聚集性的存在。因此,如何选择窗口形状,值得深入研究。

本研究从时空两个维度进行了探测,尚未考虑野生蕈中毒聚集性的人群分布特征。如果某次野

生蕈中毒的发生主要集中在某一年龄段和性别的人群中,若未将年龄和性别纳入模型就有可能探测不到该聚集性的存在。而将性别和年龄等其他人口学特征纳入模型,则成为真正意义上的多维聚集性探测。这有助于提高探测效率和分析协变量与聚集性的联系,了解野生蕈中毒聚集性的各种因素的异常分布形态如自然环境因素、社会因素以及发病人群特征等的异常分布形态以及它们与野生蕈中毒聚集性的关联。

参考文献

- [1] Kulldorff M. Scan statistics for geographical disease surveillance [M]. England: John Wiley and Sons Ltd, 2005: 115-131.
- [2] Kulldorff M. Prospective time-periodic geographical disease surveillance using a scan statistic [J]. Journal of the Royal Statistical Society, 2011, 164(1): 61-72.
- [3] 苏发旺, 白光平. 云南省姚安县 2000—2010 年食物中毒事件分析 [J]. 职业与健康, 2012, 28(8): 969-970.
- [4] 刘晓强, 秦明芳, 郑燕, 等. 2005 年云南省食物中毒分析及防治策略探讨 [J]. 卫生软科学, 2007, 21(2): 163-164.
- [5] Rogerson P A. Monitoring point patterns for the development of space-time clusters [J]. Journal of the Royal Statistical Society, 2001, 164(1): 87-96.
- [6] Goldenberg A, Shmueli G, Caruana R A, et al. Early statistical detection of anthrax outbreak by tracking over-the-counter medication sales [J]. Journal of Proceedings of the National Academy of Science, 2002, 99(8): 5237-5249.
- [7] 云南省统计局, 国家统计局云南调查总队. 云南省 2011 年统计年鉴 [M]. 北京: 中国统计出版社, 2011: 43.
- [8] Kulldorff M. A spatial scan statistic communications in statistics [J]. Journal of Theory and Methods, 1997, 26(6): 1481-1496.
- [9] Kulldorff M. Information management service, Inc. satscan user guide [Z]. 2007.
- [10] Jones R C. Use of a prospective space-time scan statistic to prioritize shigellosis case investigation in an urban jurisdiction [J]. Journal of Public Health, 2006, 121(2): 133-139.
- [11] Pearl D L. The use of outbreak information in the interpretation of clustering of reported cases of *Escherichia coli* O157 in space and time in Alberta, Canada, 2000-2002 [J]. Journal of Epidemiological Infections, 2006, 134(4): 699-711.
- [12] Lazarus R, Kleinman K, Dashevsky I, et al. Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events [J]. Journal of Emerging Infectious Diseases, 2002, 8(8): 753-760.