

Knox 方法在传染病时空聚集性探测中的应用

刘巧兰 李晓松 冯子健 马家奇

【摘要】 目的 探讨 Knox 方法在传染病时空聚集性探测中的应用。方法 介绍 Knox 方法的原理,以 2005 年 4-5 月北京市流行性脑脊髓膜炎(流脑)病例为例,介绍其统计分析过程并解释结果。结果 在选定时间临界值 13 d 和空间临界值 29.67 km 时,流脑可能存在时空聚集性。选取其他时间和空间临界值,均无时空聚集性。结论 在根据传染病的特点并选择合适的时间与空间临界值的条件下,Knox 方法能够较好地探测传染病的时空聚集性,对原因不明的疾病可以根据不同的临界时间和临界空间“诊断”其是否有传染性,因此对传染病早期预警有一定作用。

【关键词】 流行性脑脊髓膜炎; Knox 方法; 时空聚集性

Study on the application of Knox method to temporal-spatial cluster for infectious diseases LIU Qiao-lan*, LI Xiao-song, FENG Zi-jian, MA Jia-qi. *Department of Health Statistics, School of Public Health, Sichuan University, Chengdu 610041, China

【Abstract】 Objective Application of Knox method was discussed in exploring temporal-spatial cluster for infectious diseases. **Methods** The theory of Knox method was introduced in detail. As an example, the temporal-spatial cluster of epidemic cerebrospinal meningitis cases from April to May in 2005 was explored in Beijing and the statistical process of Knox method was clearly demonstrated with related results well interpreted. **Results** Only when time and distance threshold values were set as 13 days and 29.67 km, potential temporal-spatial cluster could be detected in epidemic cerebrospinal meningitis cases. **Conclusion** According to the characteristics of the epidemics, under Knox method, if appropriate threshold values of time and distance were chosen, information on temporal-spatial cluster of infectious diseases could be obtained. According to the records of different time and distance threshold values, the nature of an infectious disease might be explored. Thus Knox method could help us to strengthen the early warning system on infectious disease control strategy.

【Key words】 Epidemic cerebrospinal meningitis; Knox method; Temporal-spatial cluster

传染病时空聚集性在某种意义上是指传染病存在时间和空间的交互作用,即传染病在时间上接近,空间上也比较接近^[1,2],时空聚集性统计方法的思想其实就来自于传染病的流行病学模型,疾病如具有传染性,会致使病例在时间和空间上有交互作用。传染病时空聚集性主要用于探讨传染病病因、流行模式等,也可以用来探测疾病的流行和暴发趋势。建立和完善传染病早期预警系统(early warning system)是防控传染病的重要手段之一,而传染病的时空聚集性探测(temporal-spatial cluster detection)则是建立传染病早期预警系统的重要基

础之一^[3]。我国于 2004 年初建立全国传染病疫情网络直报系统,但常规监控(conventional surveillance)的早期预警系统还没有完全建立起来,先进的聚集性统计方法有待于进一步探索。本文以流行性脑脊髓膜炎(流脑)为例,探讨 Knox 方法在传染病时空聚集性探测中的应用价值。

基本原理

假设有 n 个病例,可配成 $N = n(n-1)/2$ 个病例对子,在同时设定空间和时间临界值后,就可以判断病例对子在空间上和时间上“远”还是“近”。然后构建检验统计量 X , X 是小于空间临界值和时间临界值的病例对子数^[1,2,4]:

$$X = \sum_{i=1}^n \sum_{j=1}^{i-1} S_{ij} t_{ij}$$

式中, S_{ij} 是空间邻接值,如果病例 i 和 j 之间的距离小于设定的空间临界值则为 1,否则为 0; t_{ij} 是时间

基金项目:国家自然科学基金资助项目(30571618);WHO 基金资助项目(P/2004/CHN/CSE/1.1/001, activity code: 03.01.04. AW. 01-02)

作者单位:610041 成都,四川大学华西公共卫生学院卫生统计学教研室(刘巧兰、李晓松);中国疾病预防控制中心疾病控制与应急处理办公室(冯子健);公共卫生监测与信息服务中心(马家奇)

邻接值,如果病例 i 和 j 之间等待时间小于设定的临界值则为 1,否则为 0。当时空交互作用存在的时候,处于临界值内的每对病例会互相靠近,统计量会变得很大。 X 的期望和方差为^[5]:

$$E[X|N_s, N_t] = N_s N_t / N$$

$$V(X|N_s, N_t, N_{2s}, N_{2t}) = \frac{N_s N_t}{N} + \frac{4N_{2s} N_{2t}}{n(n-1)(n-2)} + \frac{4[N_s(N_s-1) - N_{2s}][N_t(N_t-1) - N_{2t}]}{n(n-1)(n-2)(n-3)} - \left(\frac{N_s N_t}{N}\right)^2$$

式中, N_s 、 N_t 分别是空间和时间上距离小于界值的病例对子数; N_{2s} 、 N_{2t} 分别是空间上和时间内小于界值的病例对子的对子数。可以用蒙特卡罗 (Monte Carlo) 模拟推断和 χ^2 检验对统计量 X 进行检验。本文采用蒙特卡罗方法,即先假定病例间的时间距离和空间距离是独立的,空间距离和时间距离矩阵可以割裂,保持病例间的空间距离矩阵元素不动,随机重排时间距离矩阵元素,在随机重排数据集的基础上再重新计算统计量 X ,如果实际上存在时空交互作用,则实际计算的 X 相对于随机重排的 X 分布,会远离该分布或处于该分布的尾部,从而判断其有统计学意义^[4,6]。模型拟合主要在 ClusterSeer 2.2.6 中实现。

实例分析

1. 资料来源:流脑资料来源于中国疾病预防控制中心 2005 年全国传染病疫情网络直报系统数据库(未含港、澳、台地区数据),报告病例的信息包括诊断信息(发病时间等)、病例个体信息(性别、年龄、住址等)。本文主要以北京 2005 年 4-5 月流脑病例为例,共提取流脑数据 45 例,占北京 2005 年全年病例数的 52.94%。

地理坐标数据来源于国家基础地理信息系统 (GIS) 网站、经纬度网络查询系统、中国地图册^[7]。理论上,通过 GIS 9.0 系统读取具有详细地名的中国电子地图,可以获得每个病例的具体地理坐标(经纬度),如流脑某病例具体地址为“北京市海淀区清河街道永泰西里 × × 号”,通过 GIS 9.0 系统应该可以给出该地点经纬度坐标,精确到度、分、秒,但是如果 GIS 系统电子地图地名不能详细到某街道某号或某乡某村某组,则无法获得该病例精确的经纬度坐标。本研究只获得了最小单位为县(直辖市具体

到区)GIS 系统电子地图,结合经纬度坐标查询系统以及中国地图册,只能将地理坐标估计到乡镇或街道。假如北京市发生 1 例流脑病例的地址为“北京市丰台区卢沟桥乡 × × ×”,本研究只能估计到“卢沟桥乡”,其经纬度坐标估计为 (116.2106, 39.8624)。地址不详细的病例经纬度坐标用其所在区县的行政中心经纬度代替。图 1 显示了 45 例流脑病例的空间分布,其中黑点表示病例(为了显示清楚,病例质点被放大),如第 36 号病例发生于 2005 年 4 月 26 日,经纬度估计为 (116.536, 40.3624)。

2. Knox 方法探测结果及其解释:Knox 方法关键的步骤是确定时间和空间临界值。没有历史及其他国家数据资料作参考,则本研究以下列时间和空间临界距离作探索:流脑发病潜伏期为 1-7 d,各病例平均时间距离为 13 d,取 1-13 d 作为时间临界值范围进行探索。若临界时间设定为 3 d,即 2 个病例出现相差时间 < 3 d 为“近”, > 3 d 为“远”。由于当今交通非常发达,确定空间临界距离比较困难,各病例平均空间距离约为 29.67 km,因此,取 1-29.67 km 作为空间临界值范围进行探索。若空间临界距离设定为 20 km,意即 2 个病例空间距离 < 20 km 为“近”, > 20 km 为“远”。

当临界时间选取 13 d,临界空间选取 29.67 km 时,通过蒙特卡罗方法 999 次模拟, $P = 0.058$,接近于通常假设的检验水准 0.05,根据北京市流脑历年在 4-5 月份的实际发病情况,可以推断流脑可能有时空聚集性。图 2 和图 3 显示病例对子在单纯空间和单纯时间距离小于临界值 29.67 km 和 13 d 的分布情况,图 4 显示病例对子同时小于临界空间和时间的分布情况,图 5 显示病例对子在上述三种情况统一于一张地图的分布情况。图 6 为蒙特卡罗模拟结果, Knox 检验统计量 X 为 375,处于蒙特卡罗模拟的 X 分布的尾部,且 $P = 0.058$,提示了流脑可能存在时空聚集性。

在临界时间 13 d、临界空间 29.67 km 的基础上,增加或减少时间和空间临界值,蒙特卡罗方法推断均显示流脑不存在时空聚集性。图 7-9 显示当空间临界值取 10 km、时间临界值取 3 d 的病例对子分布情况;图 10 显示了蒙特卡罗方法推断结果 ($X = 24, P = 0.725$)。由于篇幅所限,取其他临界值的图和结果没有列出。



图1 2005 年 4 - 5 月北京市流脑个体病例空间分布



图2 病例对子空间距离小于临界值 29.67 km ($D=29.67$ km) 的分布



图3 病例对子时间距离小于临界值 13 d ($T=13$ d) 的分布

讨 论

Knox 方法在严格设定时间和空间临界值的条件下,能够对传染病时空交互作用进行数量化探讨,找出传染病在时间和空间上互为作用的变化规律^[7]。对于常规监测的传染病,其传染源与传播途径以及易感人群都比较清楚,若发现某地区常规监测的某传染病存在时空聚集性,从统计学角度提示其可能存在“聚集”,这种“聚集”可能没有流行病学意义,也可能提示该传染病正在“流行”。

如果通过相关知识和资料判断该传染病“流行”,则可以根据该传染病时空聚集性的临界值,迅速切断传播途径,相邻地区也应该迅速做出反应,做到及时扑灭疫情。对于原因不明的疾病,根据首发病例与后续病例的发病时间和距离,设定很多临界时间值和临界空间值,“诊断”其时空聚集性特点,探索聚集特点



图4 病例对子时间和空间距离均小于临界值 ($T=13$ d, $D=29.67$ km) 的分布

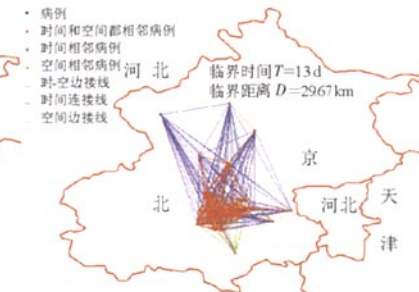


图5 病例对子时间和空间统一在一张地图上的分布

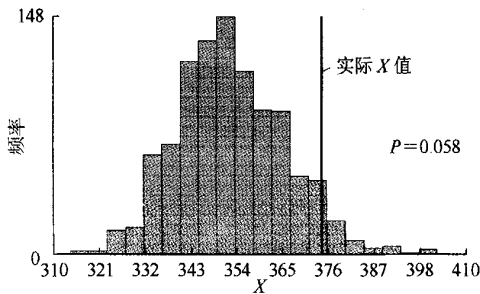


图6 Knox 检验统计量蒙特卡罗法推断结果 ($T=13$ d, $D=29.67$ km)



图7 病例对子空间距离小于临界值 10 km ($D=10$ km) 的分布



图8 病例对子时间距离小于临界值 3 d ($T=3$ d) 的分布



图9 病例对子时间和空间距离均小于临界值 ($T=3$ d, $D=10$ km) 的分布

及是否传染^[2,5,8],因此 Knox 方法对传染病早期预警有一定价值。

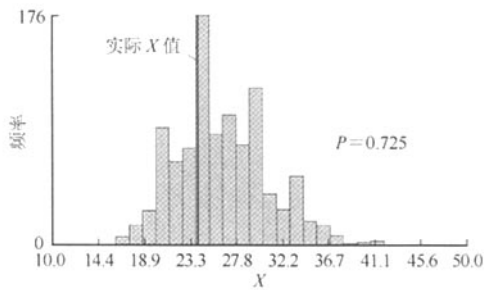


图10 Knox 检验统计量蒙特卡罗法推断结果
($T = 3 \text{ d}, D = 10 \text{ km}$)

通常情况下, Knox 方法对时间和空间临界值的选择是比较主观的,对于已知传染病,可以参考其潜伏期和传播特点来选择时间和空间临界值。一般情况下,时间和空间临界值的选择不超过所有病例的平均时间和平均空间距离^[1,8]。由于本研究无法获得每个个体病例确切的地理坐标,因此将流脑存在时空聚集性的临界时间与临界空间分别粗略地定为 13 d 和 29.67 km。该值可以作为今后北京地区流脑时空聚集性探测的参考。

Knox 方法存在不足,即暴露人口的变动,会影响其探测结果。本研究采取两种方法来解决该问题^[5]。其一,限定时间和空间范围。本研究限定为 2005 年 4-5 月北京地区,在这样的条件下,暴露人口变动不大,不会影响其结果。其二,采用蒙特卡罗方法进行统计推断,可以有效地校正暴露人口变动带来的偏性。蒙特卡罗方法是一种统计模拟方法,其基本思想基于大数定律思想,即当 n 很大时,可以用事件发生的频率来代替事件的概率。当不知道函数分布或数学结构太复杂时,在无效假设的基础上,通过函数关系或给定数学规则,通过大量的随机模拟获得统计量的分布,将实际值与该分布比较,可以做出统计推断。如果实际值位于该分布尾部或远离该分布,则可以拒绝无效假设。

当传染病潜伏期较长,发病率很高,则 Knox 方法效率不高。而流脑的潜伏期短,发病急,但发病率并不高,满足 Knox 方法的应用条件。结合 GIS^[9],Knox 方法可以清楚地描绘流脑发病的时空聚集性的特点,查找每一个病例或一群相同经纬度病例的具体地理位置(图 1),为控制其“流行”提供依据。

Knox 方法应用关键是需要知道每一个病例确切的地理坐标(经纬度),目前在实际应用中存在一定困难。但随着 GIS、全球定位系统(global positioning system)以及遥感系统(remote sensing)在传染病防控中的广泛应用,对每一个传染病病例进行定位会成为可能,因此 Knox 方法在传染病聚集性探测中具有重要的应用价值。

参 考 文 献

- [1] Knox G. The detection of space-time interactions. *Applied Statistics*, 1964, 13: 25-29.
- [2] Norstrom M, Pfeiffer DU, Jarp J. A space-time cluster investigation of an outbreak of acute respiratory disease in Norwegian cattle herds. *Preventative Veterinary Medicine*, 2000, 47: 107-119.
- [3] Michael MW, Tsui FC, Jeremy UE. The emerging science of very early detection of disease outbreaks. *J Public Health Manage Pract*, 2001, 7(6): 50-58.
- [4] Jacquez GM. A k-nearest neighbor test for space-time interaction. *Statistics in Medicine*, 1996, 15: 1935-1949.
- [5] Kulldorff M, Hjalmar U. The Knox method and other tests for space-time interaction. *Biometrics*, 1999, 55(2): 544-552.
- [6] 王岩. Monte Carlo 方法应用研究. *云南大学学报(自然科学版)*, 2006, 28(S1): 23-26.
- [7] 地质出版社地图编辑二室编. *实用中国地图册*. 北京: 地质出版社, 2006: 6-9.
- [8] Gilman EA, McNally RFQ, Cartwright RA. Space-time clustering of acute lymphoblastic leukaemia in parts of the UK (1984 - 1993). *Eur J Cancer*, 1999, 35: 91-96.
- [9] 戚晓鹏, 吕繁, 何武. 地理信息系统在流行病学中的应用及开发. *中华流行病学杂志*, 2004, 25(11): 997-999.

(收稿日期: 2007-03-19)

(本文编辑: 张林东)