

中国 2014 年结核病例流动大数据分析

史倩楠 马家奇

102206 北京, 中国疾病预防控制中心公共卫生监测与信息服务中心

通信作者: 马家奇, Email: majq@chinacdc.cn

DOI: 10.3760/cma.j.issn.0254-6450.2016.05.016

【摘要】 目的 研究 2014 年全国法定报告传染病(不含港澳台)中结核病报告病例的就诊流动情况。方法 利用 R 语言软件对 2014 年结核病报告患者进行分析及可视化, 研究结核病例就诊分布和流动情况及各省病例流出-流入-就诊医院级别情况。结果 结核病例的就诊流向在省内主要以省会城市为主, 在省间主要以从中西部省份流向东部一线北京、上海等大城市及广东地区; 流动主要以邻近省份之间居多, 同时跨省就诊病例大多就诊于三级甲等医院。结论 我国结核病流行形势依然严峻, 结核病例的跨地区流动就诊会在一定程度上促进结核的传播与流行, 需要继续深入研究病例流动就诊对结核病传播的影响, 为制订更好的结核病防控措施提供依据。

【关键词】 结核病; 病例流动; 大数据

基金项目: 国家重大科学研究计划(2012CB955504)

Big data analysis of flow of tuberculosis cases in China, 2014 Shi Qiannan, Ma Jiaqi

Public Health Surveillance and Information Service Center, Chinese Center for Disease Control and Prevention, Beijing 102206, China

Corresponding author: Ma Jiaqi, Email: majq@chinacdc.cn

【Abstract】 Objective To describe the flow of tuberculosis (TB) cases for medical care in China in 2014. **Methods** The analysis and visualization of reported flow of TB cases were conducted with software R to indicate the case distribution/flow and the levels of hospital where TB cases sought medical care in different provinces. **Results** In a province, the TB cases mainly flowed to the province capital, but the inter-provincial flow was mainly to neighbor provinces and from central or western provinces to eastern provinces (municipality), such as Beijing, Shanghai and Guangdong. The grade III (A) hospitals were the choice of most TB cases seeking medical care in other provinces. **Conclusion** The epidemic situation of TB in China is still serious, and the flow of TB cases would facilitate the spread of the disease. It is necessary to strengthen the research on influence of TB case flow for seeking medical care on the spread of TB to provide evidence for better TB prevention and control.

【Key words】 Tuberculosis; Case flow; Big data

Fund program: National Basic Research Program of China (2012CB955504)

结核病是由结核分枝杆菌感染引起的一种慢性传染性疾病, 主要经呼吸道传播。结核病是致命的传染性疾病之一, 我国患结核病人占全球 11%, 居世界第二位, 是结核病高负担国家之一^[1]。在我国法定传染病报告系统中, 肺结核报告发病人数始终位居甲乙类传染病前列, 是重要的公共卫生问题^[2]。为了解结核病病例就诊流动情况, 制订更好的结核病防治策略和合理分配卫生资源提供决策依据, 本研究利用 R 语言软件对 2014 年全国法定报告传染病中报告的结核病病例的就诊流动情况进行分析。

资料与方法

1. 资料来源: 利用中国疾病预防控制中心

法定传染病报告系统, 收集 2014 年全国按审核日期的所有结核病报告个案数据(不含港澳台)。全国省、地市、区县、乡镇地图图层(.shp)资料来源于中国 CDC。

2. 主要观察指标: 本研究主要根据个案数据中的现居住地国际编码、报告单位地区编码和医院分级分等, 对 2014 年结核病报告病例的就诊流动情况进行分析及可视化。现居住地国际编码和报告单位地区编码均为 8 位数字编码, 分别代表病例居住地和就诊地, 每两位数字分别表示省、地市、区县和乡镇, 将现居住地国际编码和报告单位地区编码匹配判断病例的就诊流动情况, 前 6 位编码相同代表在本区县内就诊, 仅前 4 位编码相同代表在本市不同

区区间流动就诊,仅前2位编码相同代表在本省不同地市间流动就诊,前2位编码不同代表跨省流动就诊。

3. 研究方法:采用R语言软件、ggplot2、circlize、rCharts等绘图软件进行数据整理分析及可视化,将2014年全国按审核日期的所有结核病报告个案数据导入R软件,经数据整理,选出所需变量,剔除已删除卡获得已终审卡。

根据结核病报告病例的现居住地国际编码和报告单位地区编码,将病例就诊流向分为4个组:区区内、市内区区间、省内地市间和跨省;通过地区编码与地图图层数据匹配获得居住地和报告地坐标,利用ggplot2软件分别做区区内病例分布图,市内区区间、省内地市间和跨省流动图,分析病例从居住地到就诊地的流动情况。

进一步分析跨省就诊病例组数据,根据现居住地国际编码和报告单位地区编码的前2位(代表省份)统计各省之间病例流动就诊的数量,计算分省病例流入流出比[(流入数-流出数)/流入流出和],流入各省病例的前3位省份及占总流入病例的比例,再根据医院分级分等变量统计流动就诊病例到各级医院就诊数量,分别利用circlize、rCharts软件做圆环图和Sankey图,分析跨省就诊病例组中各省病例流出-流入-就诊医院级别情况。运用描述性流行病学方法根据变量终审时间、年龄、性别和职业描述跨省就诊病例的时间和人群分布情况。

结 果

1. 病例流动:数据整理后获得结核病报告病例已终审个案1 023 060例,区区内、市内区区间、省内地市间和跨省就诊流向4个组分别占72%、18%、7%和3%。

2. 病例空间动态分布:区区内、市内区区间、省内地市间和跨省流动就诊情况见图1~4。从区区内和市内区区间就诊情况图中可观察到结核病高发聚集分布地区,东北主要集中在哈尔滨和长春市;华北主要集中在北京、河北;华东主要集中在长三角地区和安徽、江西省;华中主要集中在湖北、湖南、河南省;华南主要集中在广东、广西;西南主要集中在重庆、四川、贵州省;西北主要集中在新疆的喀什与和田(图1、2)。根据省内地市间就诊流向图(图3),可发现省内流动多聚集于省会等大城市;在跨省流动就诊病例中(图4),病例大多流向北京、广东、浙江、上海等东部一线发达省份。

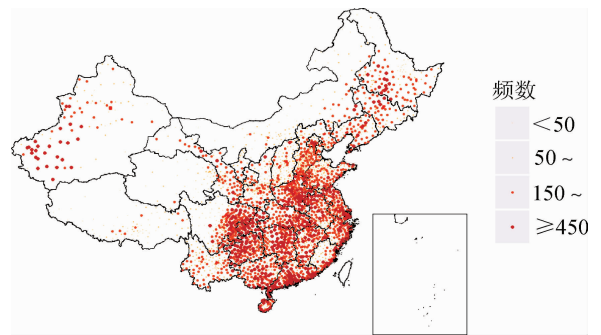


图1 2014年全国结核病报告病例区区内就诊情况

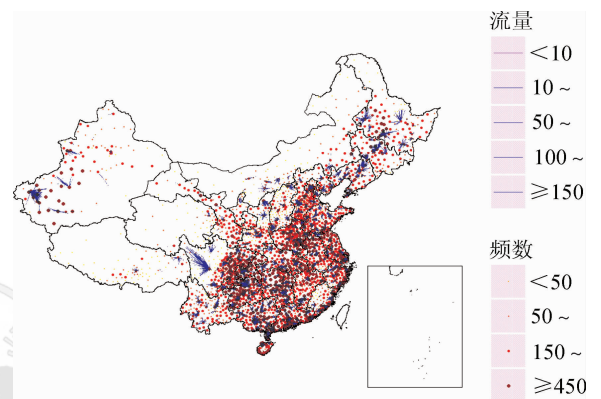


图2 2014年全国结核病报告病例市内区区间就诊情况

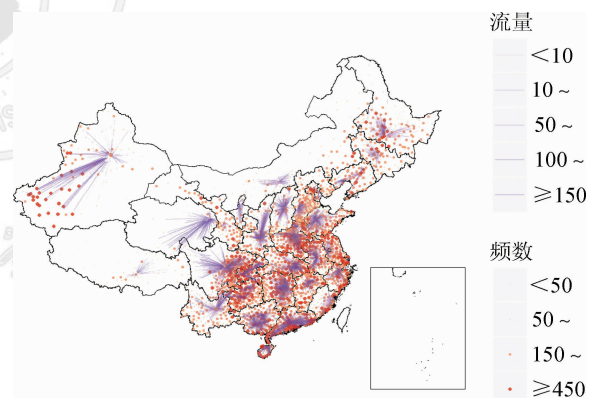


图3 2014年全国结核病报告病例省内地市间就诊情况

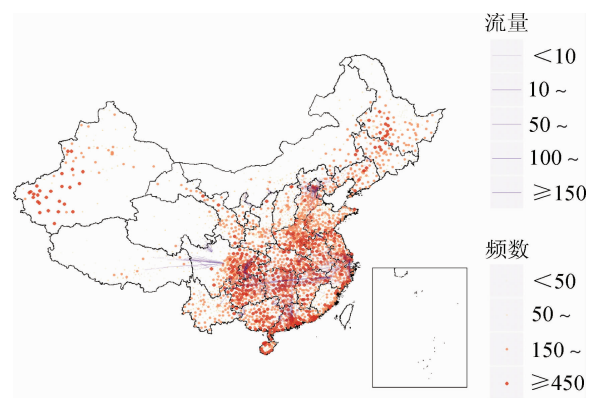


图4 2014年全国结核病报告病例跨省就诊情况

3. 病例流入流出比:各省流入流出比见图5。流入流出病例比>50%的省份主要有北京、上海、

广东、天津、浙江、宁夏和江苏,分别为98%、96%、75%、74%、62%、57%和54%。在跨省就诊病例组中,各省之间病例流入流出情况见图6。流出病例较多的前5个省份是四川、河北、贵州、安徽、河南,分别为2 868、2 847、2 833、2 133、1 935例;流入病例较多的前5个省份是北京、广东、浙江、上海、江苏,分别为5 086、4 363、3 153、2 223、2 159例。

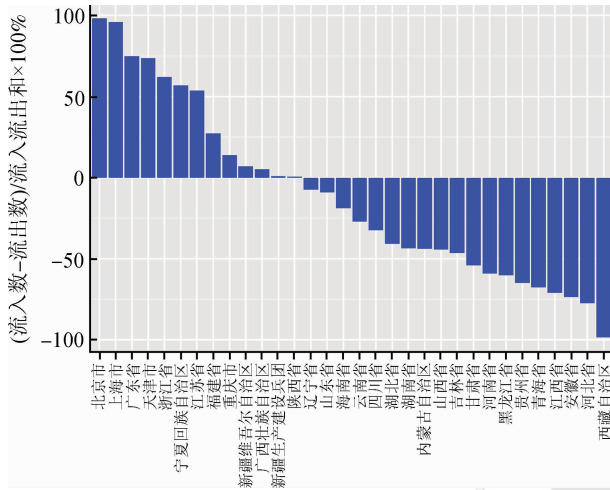


图5 2014年全国结核病报告病例跨省就诊病例各省流入流出比

4. 跨省流动病例流量与流向:各省前3位流入来源占总流入比例见图7。各省流入来源多为邻近

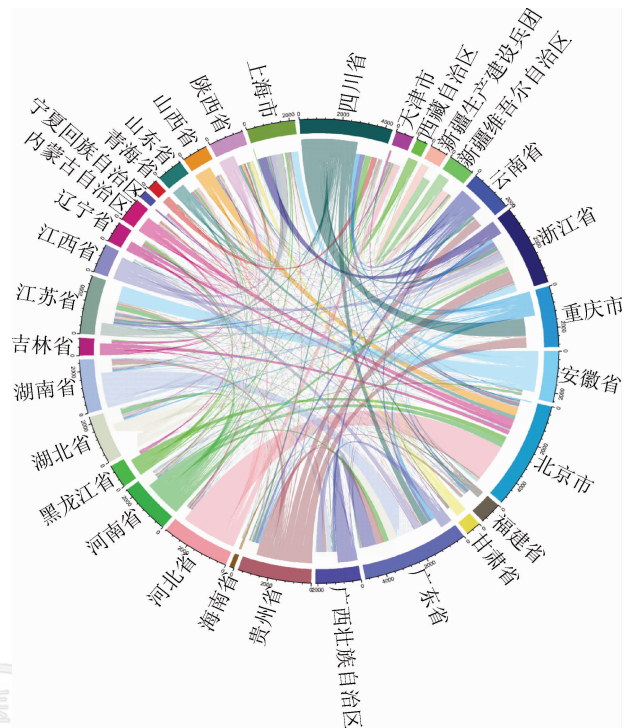


图6 2014年全国结核病报告病例跨省就诊病例分省流向圆环图

省份。北京地区流入病例主要来自河北、山西、河南,占总流入病例的55%;上海主要来自浙江、安徽、江苏,占总流入病例的56%;广东主要来自湖南、广西、湖北,占总流入病例的52%;天津主要来自河北、

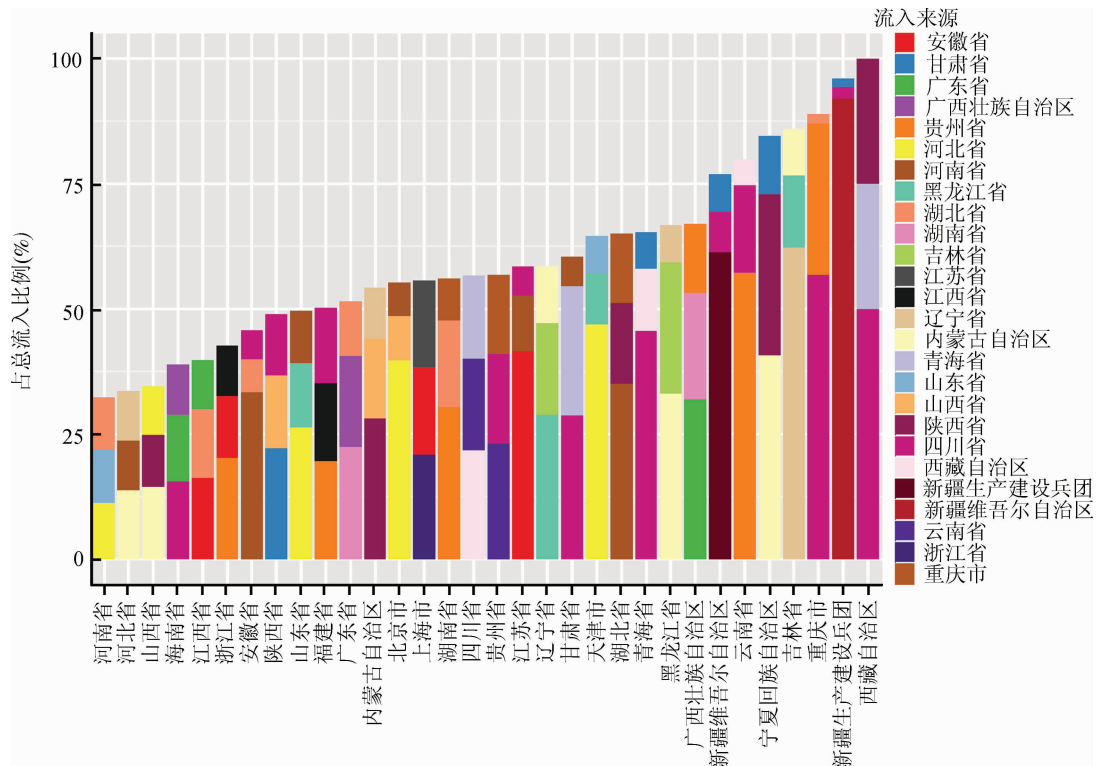


图7 2014年全国结核病报告病例跨省就诊病例各省前3位流入来源占总流入病例比例

黑龙江、山东, 占总流入病例的65%; 浙江主要来自贵州、安徽、江西, 占总流入病例的43%; 宁夏主要来自内蒙古、陕西、甘肃, 占总流入病例的85%; 江苏主要来自安徽、河南、四川, 占总流入病例的59%。

5. 跨省流动病例就诊分布: 各省流入的病例就诊医院级别情况见图8。流入病例就诊最多的医院

级别是三级甲等, 占53%, 其次为二级甲等、无级别, 分别占18%、10%。跨省流动就诊病例从居住地到报告地再到就诊医院的情况见图9。

6. 跨省流动病例流行病学特征: 在时间分布上, 各月流动就诊数差别不大, 没有明显的高峰期, 相对来说1、2月数量较少(图10)。在人群分布中, 流动就诊数性别比为2.12:1, 男性明显高于女性; 在不同年龄组中, 20~30岁青壮年人群流动就诊数最高(图11); 职业分布中, 排在前5位的是农民, 家政、家务及待业, 工人, 不详, 学生, 其中农民例数(42.74%) 远高于其他职业(图12)。

讨 论

结核病报告病例空间分布呈地区聚集性高发, 中东部地区报告数较西部地区高。在所有报告病例中, 2/3以上的病例在本区县内就诊, 但仍有部分病例省内地市间流动就诊和跨省流动就诊。李涛和刘宇红^[3]研究, 发现流动人口的增长对结核病的分布有着巨大影响。结核病例的流向在省内主要以省会城市为主, 在省间主要从中西部省份流向北京、上海、广东等东部一线发达省份, 流动主要以邻近省份之间居多。同时跨省就诊病例大多就诊于三级甲等

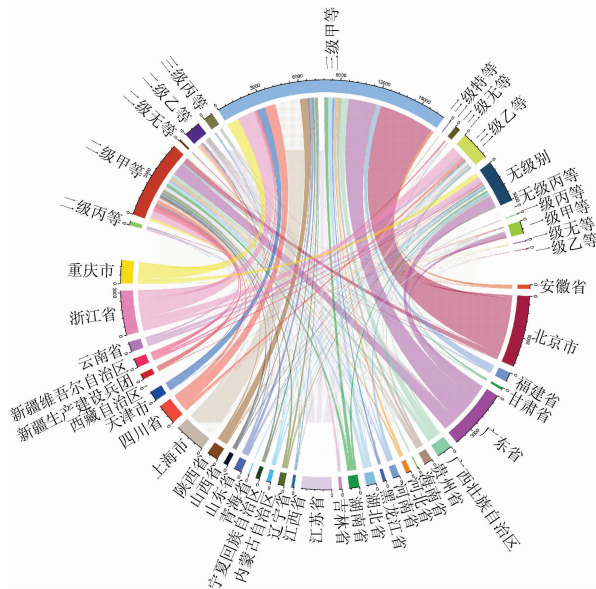


图8 2014年全国结核病报告病例跨省就诊病例就诊医院级别情况圆环图

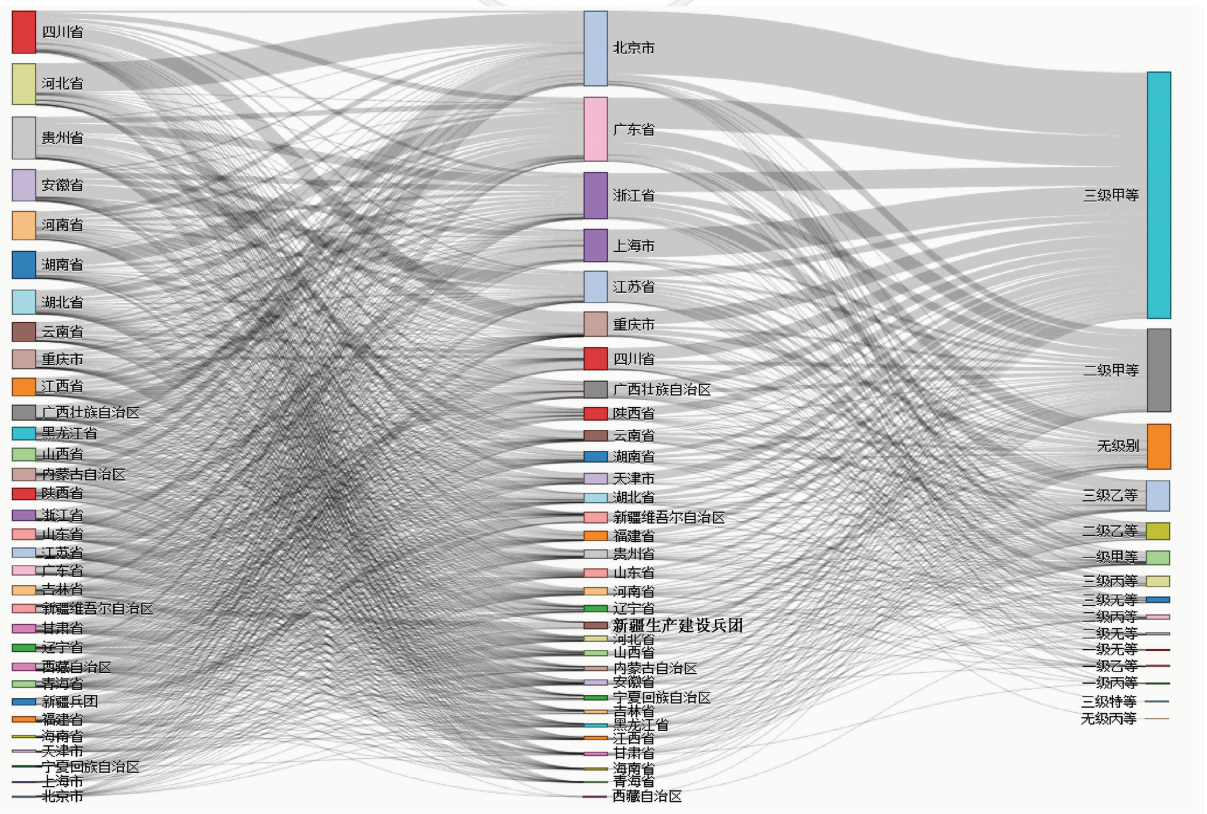


图9 2014年全国结核病报告病例跨省就诊流出-流入-就诊医院级别情况(Sankey图)

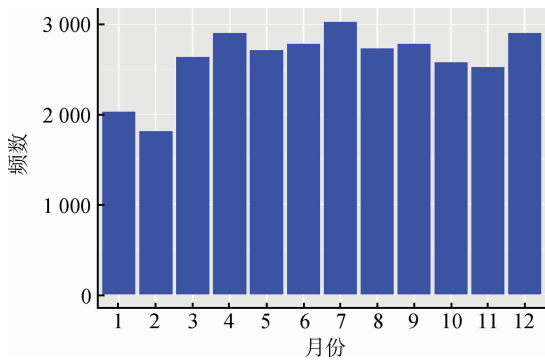


图10 2014年全国结核病报告病例跨省就诊时间分布

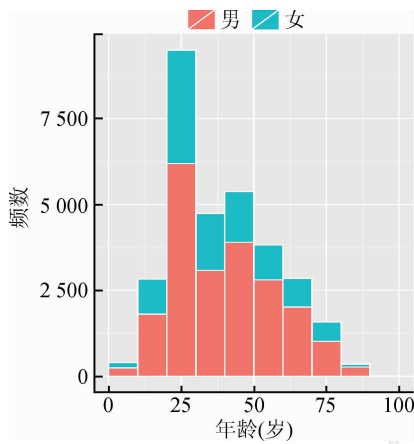


图11 2014年全国结核病报告病例跨省就诊性别年龄分布

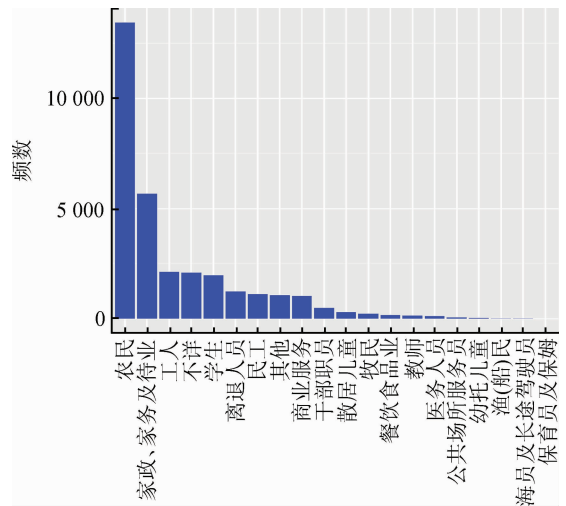


图12 2014年全国结核病报告病例跨省就诊职业分布

病例的监管,是该地区重点加强的工作;针对结核病例流动主要集中到三级甲等以上医院就诊的特点,应进一步提高三级甲等以上医院结核病诊断治疗能力,同时还须加强病例流动可能存在不规范治疗而产生耐药性的监测。

我国结核病流行形势依然严峻,需不断加强病例发现及治疗管理工作,深入研究病例流动就诊对结核病传播的影响,为制定更好的结核病防控措施提供依据。

利益冲突 无

参 考 文 献

[1] World Health Organization. Global tuberculosis report 2014[R]. Geneva: World Health Organization, 2014.

[2] 张慧,姜世闻,王黎霞. 我国结核病防治工作形势分析[J]. 中国医学科学院学报, 2009, 31(4): 393-395. DOI: 10.3881/j.issn.1000-503X.2009.04.002.

Zhang H, Jiang SW, Wang LX. Prevention and control of tuberculosis in China[J]. Acta Acad Med Sin, 2009, 31(4): 393-395. DOI: 10.3881/j.issn.1000-503X.2009.04.002.

[3] 李涛,刘宇红. 新流动人口对北京市结核病空间分布的影响[J]. 国际结核病与肺部疾病杂志, 2011, 6(2): 63-69.

Li T, Liu YH. Impact of new migrant populations on the spatial distribution of tuberculosis in Beijing[J]. Int J Tuberc Lung Dis, 2011, 6(2): 63-69.

(收稿日期:2015-12-25)

(本文编辑:王岚)

医院,人员密集和易感性较高使得结核更易传播。从人口学特征分析,其跨省流动病例中,男性明显高于女性,且以20~30岁青壮年为主,职业分布中,农民的数量远高于其他职业。在时间序列分布上,相对1、2月流动病例数量较少,可能与春节返乡过年有关。结核病是以空气传播为主要途径,所以病例就诊的流动将促进结核病的地区间传播,带来结核病暴发流行的隐患。但还需进一步分析流动就诊病例中活动性肺结核的比例,估计其对流入地疫情的影响。

通过对大数据的分析,揭示我国结核病例近30%流动就诊,结核病的流行与人口流动相关,在一定程度上促进传染病的传播与流行。结核病的流行在空间上存在明显的聚集性高发区域,因此须研究制订更具有针对性的防控策略,加强对高发地区流动人口的防控干预;北京、上海、广东、天津、浙江、宁夏和江苏是主要的病例跨省流入地区,对流动