·专题报道·

中国空气污染对人群健康影响综合信息平台设计及构建*

郝舒欣,刘悦,刘婕,吕祎然,徐东群

摘　要:目的　建立空气污染对人群健康影响综合信息平台,整合全国范围内已经建立的各种监测网的数据资源,形成空气污染及相关健康综合数据库。**方法**　2014—2015 年通过数据收集、数据统计和系统设置 3 大模块,实现数据的整合、分析和管理功能。**结果**　信息平台于 2015 年建立,目前应用于全国 10 个省 12 个城市,将环境因素数据、死因监测数据、疾病与症状监测数据以及健康调查数据有效整合,形成从数据收集、校验、审核、清理到统计分析的规范流程。**结论**　此平台的建立不仅可以实现数据的共享和综合利用,同时还能为空气污染对人群健康影响的科学研究提供数据支撑。

关键词:空气污染;$PM_{2.5}$;健康影响;综合信息平台

中图分类号:R 122.7　**文献标志码:**A　**文章编号:**1001-0580(2017)01-0009-05　DOI:10.11847/zgggws2017-33-01-03

Design and constitution of comprehensive information platform for health effects of air pollution in China

HAO Shu-xin, LIU Yue, LIU Jie, et al (*Institute of Environmental Health and Related Product Safety, China Center for Disease Control and Prevention, Beijing* 100021, China)

Abstract: Objective　To construct a comprehensive information platform for health effects of air pollution which could integrate nationwide data from various established monitoring systems to form a database on air pollution-related health. **Methods**　Three modules including data collection, statistics, and system setting were designed and established for data integration, data analysis and system management between 2014 and 2015. **Results**　The platform was established in 2015 and applied in 12 cities of 10 provinces currently. Data on environment monitoring, death registry, and disease and symptom surveillance could be integrated and standardized procedures in data collection, cleaning, verification, and statistical analysis were achieved via the platform. **Conclusion**　The established platform could not only realize data sharing and comprehensive utilization but also support researches on air pollution and health effects with its integrated database.

Key words: air pollution; particulate matter ≤2.5 μm in aerodynamic diameter; health effect; comprehensive information platform

空气污染对人群健康的影响涵盖从各种生理生化指标的变化、轻微症状的出现到患病就诊、住院、甚至死亡,并且随着污染物浓度的升高,人群出现生理反应异常、患病、死亡等各种健康效应的风险也会随之增大[1-2],世界卫生组织(World Health Organization, WHO)明确提出空气污染物尤其是颗粒物是呼吸系统、循环系统发病率和死亡率升高的重要影响因素[3]。而儿童、老年人或患有基础性疾病患者等敏感人群这类风险会更为明显[4-6]。但国内现有研究大多关注的健康效应指标各不相同[7-8],在全国范围内建立符合中国人群特征的空气污染健康影响暴露-反应关系鲜有报道[9]。本研究从中国环境健康领域的重大需求入手,于 2014—2015 年构建空气污染对人群健康影响综合监测信息平台,实现环境、气象、健康等监测数据的连接和综合利用。

1　平台设计

1.1　平台功能设计(图 1、表 1)　空气污染对人群健康影响综合信息平台可分为数据收集、数据统计和系统设置 3 大模块。其中数据收集模块按照监测目的来划分,又可细分为空气污染物及气象监测数据库、人口学及疾病监测数据库和空气污染人群健康影响调查及症状监测数据库 3 个子模块;数据统计模块则按照数据收集模块的内容分别生成相对应的数据收集情况统计报表和简单的数据计算统计报表。可根据地区、时间、指标等不同维度定制,实时呈现各个数据指标的数据分布情况或趋势图、饼图等信息;系统设置模块包含用户管理、基本数据字典管理等功能,主要用于管理用户基本信息和各城市行政区划信息以及所辖范围内的医院、学校、社区等基本信息字典(见图 1)。

* **基金项目:** 2014 年度公益性行业科研专项(201402022)
作者单位: 中国疾病预防控制中心环境与健康相关产品安全所,北京 100021
作者简介: 郝舒欣(1984-),女,山西吕梁人,助理研究员,硕士,研究方向:信息技术在环境卫生领域的应用。
通讯作者: 徐东群,E-mail:dongqunxu@126.com
数字出版日期:2016-7-19 9:15
数字出版网址:http://www.cnki.net/kcms/detail/21.1234.R.20160719.0915.004.html

数据收集模块是信息平台的核心部分，其中空气污染物及气象监测数据库主要收集气象监测数据、环境空气质量监测数据以及特别需要关注的空气污染物如细颗粒物（$PM_{2.5}$）及其成分监测数据、挥发性有机物（volatile organic compounds，VOCs）、半挥发性有机物（semi-volatile organic compounds，SVOC）等。用来分析不同地区的空气污染特征，包括不同城市的空气污染物区域差异以及月度、季度和年度变化趋势；气象条件变化对空气污染的影响以及同一个城市中不同污染条件下空气污染物特征及成分差异。人口学及疾病监测数据库主要收集目前已经在其他信息系统中运行成熟且与空气污染相关的健康监测数据，如死因监测数据、急救中心接诊数据、医院门诊数据、急诊数据及住院数据等。这一部分主要针对空气污染对人群健康影响中死亡和患病就诊这两类效应终点，分析不同城市的空气污染物暴露对监测人群总死亡以及分病种超额死亡的影响、对监测人群急救就诊的影响和对医院门诊、住院等的影响。空气污染人群健康影响调查及症状监测数据库主要收集敏感人群的健康调查数据，如小学生的症状监测数据、肺功能指标、老年人、慢性阻塞性肺疾病（chronic obstructive pulmonary disease，COPD）患者、心血管病患者或哮喘儿童的相关症状和疾病健康问卷调查数据等。这一部分主要针对空气污染对人群健康影响中症状或生理生化指标改变这一效应终点，分析不同污染条件下对其相关症状、疾病情况以及肺功能等的影响（见表 1）。

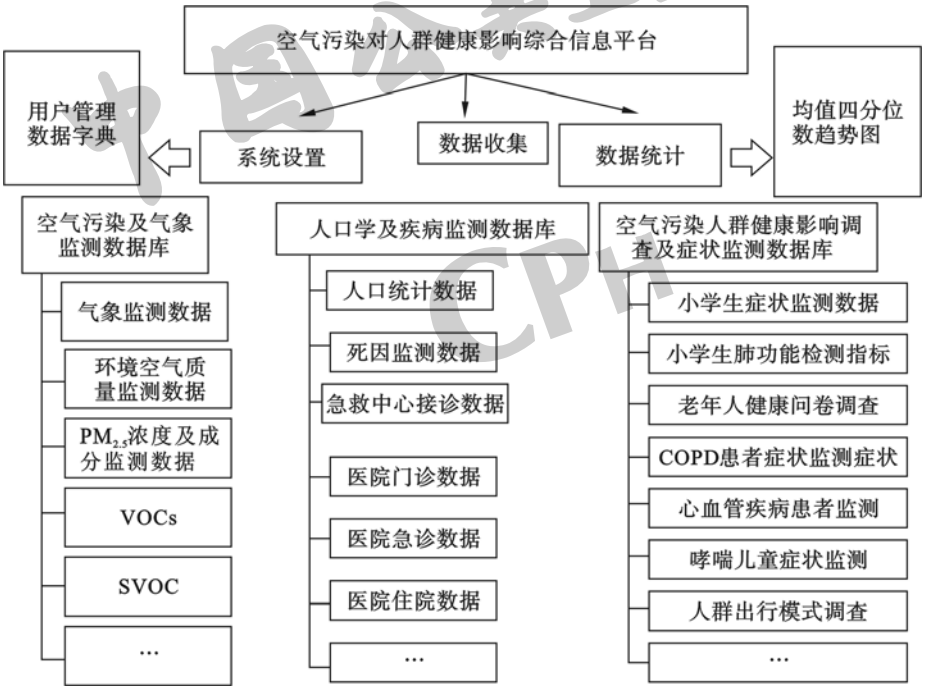

图 1　空气污染对人群健康影响综合信息平台功能图

表 1　数据收集模块各部分数据类型及指标一览表

子模块名称	数据类型	数据指标
空气污染物及气象监测数据库	气象监测数据	日平均温度、平均相对湿度、风速、降水量等指标；
	环境空气质量监测数据	二氧化硫（SO_2）、二氧化氮（NO_2）、可吸入颗粒物（PM_{10}）、一氧化碳（CO）、臭氧（O_3）、细颗粒物（$PM_{2.5}$）等指标；
	$PM_{2.5}$浓度及成分监测数据	重金属类、多环芳烃类、阴阳离子类化合物等指标；
	VOCs	苯、甲苯、二甲苯等指标；
	SVOC	邻苯二甲酸酯类、多氯联苯类化合物等指标；
人口学及疾病监测数据库	人口统计数据	户籍人口数据或常住人口数据，包括性别、年龄段分组等；
	死因监测数据	性别、年龄、死亡时间、根本死亡原因及根本死因国际疾病分类（International Classification of Diseases，ICD）编码等死因个案数据；
	急救中心接诊数据	接诊时间、患者性别、年龄、患者主要症状及初步诊断等个案资料；
	医院门诊数据	接诊时间、患者性别、年龄、主要诊断、诊断 ICD 编码等个案资料；
	医院急诊数据	接诊时间、患者性别、年龄、主要诊断、诊断 ICD 编码等个案资料；

续表 1　数据收集模块各部分数据类型及指标一览表

子模块名称	数据类型	数据指标
空气污染人群健康影响调查及症状监测数据库	医院住院数据	入院时间、出院时间、患者性别、年龄、主要诊断、诊断 ICD 编码等个案资料;
	小学生症状监测数据	监测日期、姓名、性别、年龄、天气状况、缺课情况、症状等指标;
	小学生肺功能检测指标	检测时间、性别、年龄、身高、体重、用力肺活量(forced vital capacity, FVC)、1 秒用力呼气量(forced expiratory volume in one second, FEV1)、呼气峰值流速(peak expiratory flow, PEF)等指标;
	老年人健康问卷调查	居住环境调查及疾病和症状问卷调查
	COPD 患者症状监测	监测日期、姓名、性别、年龄、天气状况、现患疾病情况、症状等指标;
	心血管疾病患者症状监测	监测日期、姓名、性别、年龄、天气状况、现患疾病情况、症状等指标;
	哮喘儿童症状监测	监测日期、姓名、性别、年龄、天气状况、现患疾病情况、症状等指标;
	人群出行模式调查	不同类型人群日常出行模式问卷调查

1.2　平台流程设计(图 2)　空气污染对人群健康影响综合信息平台的工作流程主要包括两个方面:系统配置流程和数据导入审核流程。系统配置流程主要由管理员负责,用于完成系统的初始化设置,包括用户的创建、角色的配置、基本数据字典的配置和管理等;数据导入及审核流程主要由录入员和审核员负责,用于完成数据的导入、查询、修改、提交、审核等。其中数据的审核分为两级,分别由地市级负责数据的一级审核,省级负责数据的二级审核,数据状态分为未提交、一级审核通过、二级审核通过 3 种情况,数据仅在未提交状态下才可进行删除和修改。考虑到气象、环境空气质量、死因、急救资料等均存在数据量大的问题,无法做到人工逐一进行核实检查,为方便用户,平台在数据导入前设置了对数据进行自动校验和批量审核的功能,同时为保证数据质量,方便国家级审核员及时反馈问题数据,平台设置了审核报告提交模块,国家级审核员可将问题数据以报告形式反馈至省级地方管理员处,由地方用户负责数据的修改和重新上报,并仅为国家级系统管理员设有特别的权限可以将已审核的数据退回重新修改。

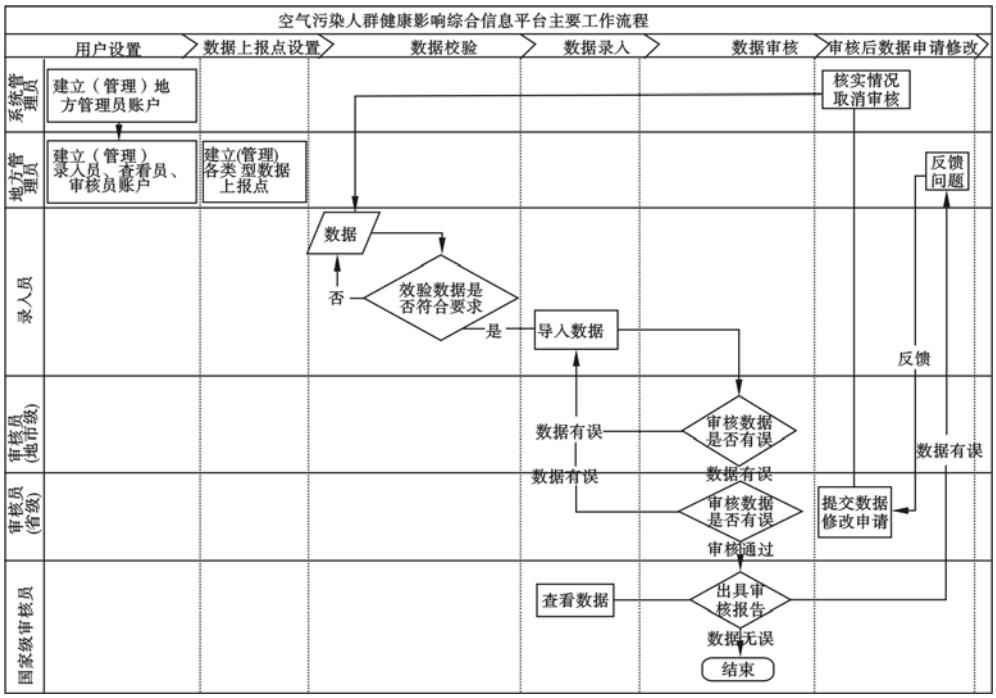

图 2　空气污染对人群健康影响综合信息平台主要工作流程图

2　系统实现

2.1　数据收集　(1)数据类型:本信息平台包括两种类型的数据,第一是收集类数据,包括气象监测数据、环境空气质量监测数据、人口统计数据、死因监测数据、急救接诊数据、医院门诊数据、医院急诊数据、医院住院数据等,这些数据分散在其他监测或信息系统中,可以通过一定的方式直接获取后导入;第二是监测获取的数据,包括 $PM_{2.5}$ 浓度及成分监测数据、VOCs 监测数据、SVOC 监测数据、小学生症

状监测数据、老年人健康调查、心血管疾病患者症状监测等,需要通过开展监测或调查获取。(2)实施单位:本信息平台的构建,依托卫生行业专项课题,各课题参加单位已利用该平台实现两种不同类型数据的收集、校验、审核、清理、查询、统计等功能,该信息平台成熟后,可在全国范围的卫生系统应用,不同机构均可根据职责分工,进行用户角色的权限配置,实现相应的功能。(3)数据来源及收集方法:针对不同的数据类型,数据获取方式也各不相同。收集类型的数据,如气象监测数据来自于气象部门、环境空气质量监测数据来自于环保部门、人口统计数据和死因监测数据来自于疾病预防控制机构、急救接诊数据和医院数据则来自于急救中心和医院。监测类型的数据则由开展监测或调查单位产生,一般为各级疾病预防控制机构。(4)收集频率:根据数据类型的不同,收集数据的频率也各不相同。气象数据、环境空气质量监测数据、死因监测数据、急救接诊数据和医院数据等按日收集,人口统计数据按年度收集;监测类型的数据则根据监测方案的要求按月或按季度收集。

2.2　系统设置　空气污染对人群健康影响综合信息平台涉及的内容多、范围广、数据量大。因此为满足全国范围内监测数据的收集、管理和分析工作,平台在设计时考虑到以下几方面的要素。(1)系统安全:信息平台应从数据的存储和传输两方面保障系统安全,在数据存储方面采取双机热备和定时备份机制,确保信息平台的正常不间断运作;在应用层面和数据库层面均采用多机机制,避免单点故障。数据传输方面则应采用数字证书用户实名登陆和 SSL 链路加密方式,保障数据的传输安全。(2)系统结构:采用集中式数据管理方式,由国家级单位负责数据中心的建立;形成国家、省级两级管理机制;国家级系统管理员负责全国数据的管理和信息平台的运行维护工作,省级作为第二级信息管理中心,负责省级数据的管理和平台维护,包括用户、数据字典的创建和管理等。(3)权限管理:采用灵活的权限配置方式,省级地方管理员可按照用户的角色、负责工作内容等为用户做功能分配。(4)数据输入:本信息平台由于涉及到气象、环保、急救中心接诊、死因监测等数据量较大且已经相对成熟的监测数据,所以一般通过制定标准模板,将数据批量导入到信息平台中。模板通常为用户常用软件格式,如 Excel、Epi Data 等。(5)数据审核:无论是环境数据还是健康数据,大部分数据的年均数据量均在千、万甚至百万级别。因此,如此大的数据量需将人工审核与计算机智能审核相结合,在信息平台中建立数据校验和审核规则集,对于一些明确的规则,如缺失率、异常值以及简单的逻辑判断等可以在数据导入前或出具审核报告时给予标示。(6)数据输出与统计:在信息平台中用户可根据查询条件和个人权限自动定制统计报表或趋势图,如各指标均值、标准差,或中位数、四分位数等,以了解数据的分布情况、变化趋势等基本信息。同时考虑到环境健康研究的特点,按照用户的要求将环境数据与健康数据匹配输出,做到与专业统计软件如 SAS、R 软件等的无缝链接,方便用户进行更深层次的数据分析。如利用气象数据、环境空气质量监测数据、人口统计数据、死因监测数据或医院数据进行时间序列分析,了解空气污染对人群超额死亡或超额就诊等的影响。

3　运行现状

该信息平台于 2015 年正式投入使用,目前在卫生行业专项参加单位所在的 10 个省 12 个城市中应用,实现 12 个监测城市的气象数据、环境空气质量监测数据、$PM_{2.5}$ 浓度及成分监测数据、VOCs 监测数据、SVOC 监测数据,以及与之相对应的疾病监测数据和症状监测数据的收集、校验、审核、清理、查询等。目前监测数据达到几十万条,平台用户超过 200 多人。本信息平台现处于运行初期,虽然有效的将环境数据和健康数据进行了整合和链接,但仍然存在数据标准不统一,数据质量不高的问题,这也是本项目的一个重要研究领域。除在信息平台中制定规范的审核流程和审核规则外,更需要在数据进入信息平台之前进行有效的数据清理[10-11],以便在全国层面上建立统一的标准。随着数据的不断积累,各参与单位借助该信息平台,可结合分析目的进行空气污染对健康影响的分析,从而揭示出空气污染对人群的急性影响及长期变化趋势。

4　讨　论

美国疾病预防控制中心自 2002 年开展了国家环境健康追踪项目(Environmental Public Health Tracking, EPHT),其目的就在于支持全国性的环境健康影响研究,该项目持续性的收集环境监测数据与健康监测数据,并实现数据的整合、连接、分析和相关的健康教育工作[12]。以该项目中 $PM_{2.5}$ 和 O_3 对人群健康影响的暴露反应关系研究为例,其健康指标选择了住院数据,包含病人入院的日期、时间、主要诊断、第二诊断、地址、出生日期、性别等指标;而环境监测数据则使用了 $PM_{2.5}$ 和 O_3 的日均浓度;同时为更好的了解空气污染对呼吸系统和循环系统住院病人的影响,项目还将时间和地理信息数据作为连接指标用于分析不同时间尺度和地理纬度的暴露反应关系[13]。该项目涉及的监测数据由美国环

保署(Environmental Protection Agency, EPA)和美国疾控中心,通过签署协议的方式实现共享[14]。欧洲在没有开展基于多中心的空气污染对人群健康影响综合监测之前,空气污染对人群健康影响的研究也均是相互独立的,数据来源和研究方法各不相同,研究结果之间也无法进行比较。因此基于欧洲 26 个城市的欧洲空气污染与健康信息系统项目(Air Pollution and Health - A European Information System, APHEIS)应运而生。该项目的目标就在于建立一个合作性的综合网络平台,制定标准化的数据收集方式和分析方法,以确保各地使用统一的分析方法进行健康影响评估,为政策决策者呈现可靠的研究成果[15]。

中国目前也存在着同样的问题,研究成果大多基于个别城市[16-19],研究周期较短,数据收集范围和标准均不相同,无法得出代表全国或区域范围的空气污染对人群健康影响结果。因此本信息平台设计的目的就在于将环境数据、死因监测数据、疾病与症状监测数据以及健康调查数据等进行有效整合,形成从数据收集、校验、审核、清理到统计分析的规范流程,并建立完善的空气污染对人群健康影响基础数据库。该综合信息平台的优势还在于将急性健康影响研究指标和慢性健康影响研究指标同时纳入,即综合利用死因监测数据,疾病与症状监测数据,气象、环境空气质量监测数据等进行空气污染对健康影响的描述性分析、相关性分析、时间序列分析等,从而揭示出空气污染对人群的急性影响[20],同时利用综合信息平台积累长期监测信息,用于慢性影响的研究,填补国内综合信息利用的短板。在数据的标准化方面,借鉴国外运行成熟的监测项目,在信息平台中对数据质量提出严格的审核要求,不断完善数据校验规则,对各模块数据均建立起具有明确标准的规则库,不断细化各类指标的参考值范围以及指标之间的逻辑关系等。同时规范数据审核流程,针对各级用户明确审核职责,细化审核指标,在指标的完整性、有效性等方面设定详细的标准。

空气污染对人群健康影响综合信息平台不同于以病例上报为基础的传染病报告系统,具有结构复杂、数据类型多、数据量大等特点。在信息平台的建立初期,首先应以方便用户上报、审核、查询数据,数据输出和统计为基本理念;同时在信息平台运行和管理方面,应采用集中式数据管理方式,以便于应对平台建立初期数据结构和数据指标频繁调整的可能。但在信息平台运行平稳、业务发展成熟以及数据协调和共享机制完善后,应逐步向分布式管理方向发展,由各省建立自己的数据分中心,国家级数据中心仅需提取各类汇总数据。同时,如气象、环境空气质量等监测数据,可通过接口程序直接调用数据,提高工作效率和数据准确性。中国目前在没有形成完善的数据共享机制的情况下,利用综合信息平台,推动数据的共享和利用不失为一种很好的促进方式。

参考文献

[1] 王秦,李湉湉,陈晨,等. 我国雾霾天气 $PM_{2.5}$ 污染特征及其对人群健康的影响[J]. 中华医学杂志,2013,93(34):2691-2694.

[2] 李湉湉,崔亮亮,陈晨,等. 北京市 2013 年 1 月雾霾天气事件中 $PM_{2.5}$ 相关人群超额死亡风险评估[J]. 疾病监测,2015,30(8):668-671.

[3] World Health Organization. Review of evidence on health aspects of air pollution - REVIHAAP project - first results[Z]. Denmark: World Health Organization Regional Office for Europe, 2013.

[4] Atkinson RW, Kang S, Andenon HR, et al. Epidemiological time series studies of $PM_{2.5}$ and daily mortality and hospital admissions: a systematic review and meta-analysis[J]. Thorax, 2014, 69(7):660-665.

[5] Bowatte G, Lodge C, Lowe AJ, et al. The influence of childhood traffic-related air pollution exposure on asthma, allergy and sensitization: a systematic review and a meta-analysis of birth cohort studies[J]. Allergy, 2015, 70(3):245-256.

[6] Chen CH, Chan CC, Chen BY, et al. Effects of particulate air pollution and ozone on lung function in non-asthmatic children[J]. Envion Res, 2014, 137:40-48.

[7] 阚海东,陈秉衡. 我国部分城市大气污染对健康影响的研究 10 年回顾[J]. 中华预防医学杂志,2002,36(1):59-61.

[8] 赵珂,曹军骥,文湘阁,等. 西安市大气 $PM_{2.5}$ 污染与城区居民死亡率的关系[J]. 预防医学情报杂志,2011,27:257-264.

[9] 刘昌景,黄飞,杨志洲,等. 我国空气污染物与人群呼吸系统疾病死亡急性效应的 META 分析[J]. 中华流行病学杂志,2015,36(8):889-895.

[10] 刘悦,郝舒欣,韩京秀,等. 门诊个案数据快速清理及诊断疾病自动编码方法研究[J]. 中国医院管理,2015,35(9):69-71.

[11] 杜艳君,陈晨,李湉湉. 大规模数据清理的相关问题与探讨[J]. 环境与健康杂志,2014,31(4):353-354.

[12] Matte TD, Cohen A, Dimmick F, et al. Summary of the workshop on methodologies for environmental public health tracking of air pollution effects[J]. Air Quality Atmosphere and Health, 2009, 2(4):177-184.

[13] Talbot TO, Haley VB, Dimmick WF, et al. Developing consistent data and methods to measure the public health impacts of ambient air quality for Environmental Public Health Tracking: progress to date and future directions[J]. Air Quality Atmosphere and Health, 2009, 2(4):199-206.

[14] Parker JD, Woodruff TJ, Akinbami LJ, et al. Linkage of the US National Health Interview Survey to air monitoring data: an evaluation of different strategies[J]. Environmental Research, 2008, 106(3):384-392.

[15] Medina S, Tertre A, Saklad M. The Apheis project: Air Pollution and Health-A European Information System[J]. Air Quality Atmosphere & Health, 2009, 2(4):185-198.

[16] 谢鹏,刘晓云,刘兆荣,等. 我国人群大气颗粒物污染暴露-反应关系的研究[J]. 中国环境科学,2009,29:1034-1040.

[17] Kan HD, Stephanie JL, Chen GH, et al. Season, sex, age, and education as modifiers of the effects of outdoor air pollution on daily mortality in Shanghai, China: the public health and air pollution in Asia(PAPA) Study[J]. Environ Health Perspect, 2008, 116(9):1183-1188.

[18] 丁玲,朱道娟,彭东红. 大气颗粒物与儿童哮喘住院人数相关性的 Meta 分析[J]. 中华儿科杂志,2015,53(2):129-135.

[19] 黄晓亮,戴灵真,卢萍,等. 广州市 2004—2008 年大气污染对城区居民每 Et 死亡率影响的时间序列分析[J]. 中华流行病学杂志,2012,33(2):210-214.

[20] 吴伟栋,应晓华. 应对公共健康问题的症状监测[J]. 中国卫生事业管理,2015,32(6):477-479.

收稿日期: 2016-01-05　　(吴少慧编辑　解学魁校对)