

国家卫生计生监督大数据应用实践—— 非法行医专题分析

王 晖

(国家卫生计生委卫生和计划生育监督中心 北京 100007)

〔摘要〕 从卫生计生监督数据应用需求出发,介绍全国卫生计生监督大数据特征及应用架构,应用大数据技术对全国非法行医案件高发地区进行分析,阐述非法行医相关影响因素,为降低全国非法行医案发率提供应用策略。

〔关键词〕 非法行医;大数据;卫生计生监督

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2017.01.014

National Health and Family Planning Supervision Big Data Application Practice——thematic analysis of illegal medical practice

WANG Hui, *The Center of Inspection and Supervision, National Health and Family Planning Commission (NHFPC), Beijing 100007, China*

〔Abstract〕 Based on the data application requirements of health and family planning supervision, the paper introduces the big data characteristics and application architecture of national health and family planning supervision, analyzes the regions with frequently occurred illegal medical practice cases nationwide by making use of big data technology, describes relevant influencing factors of illegal medical practice, and provides application strategy for reducing occurrence rate of illegal medical practice cases nationwide.

〔Keywords〕 Illegal medical practice; Big data; Health and family planning supervision

1 引言

受利益的驱使,近年来非法的无证、无照私人诊所日益增多,非法行医中致死致伤事件也呈高发趋势,打击非法行医工作面临严峻形势。国家卫生计生委卫生和计划生育监督中心(以下简称“监督中心”)通过5年的信息化建设,构建了覆盖全国范围的信息报告系统,收集了全国所有医疗机构的存档信息、监督检查以及案件查处信息,为全国卫生计生监督业务的开展提供了基础。目前,监督中心在梳理全国卫生计生监督数据资源框架的基础

上,逐步探索应用大数据技术,通过搭建国家卫生计生监督大数据平台,利用大数据技术实现对多源数据的关联性、多维度、相关因素的深入分析,提供灵活直观的展现。本着“聚合数据、探索关系、发挥价值”的原则,关注影响人民健康的重点领域,归集与非法行医查处、公共场所监管和生活饮用水监管等方面的监督管理数据,同时聚合各地经济收入、人均医疗服务人员等相关数据,探索管理相对人、区域分布、时间趋势、医疗资源配置等因素与非法行医等不良从业行为之间的关系,提高卫生计生监督各业务、各专业数据之间的关联关系分析能力、趋势分析以及预测分析能力,更好地辅助卫生计生监督管理决策,有针对性地指导各地卫生计生监督工作,是卫生计生领域大数据应用的一个有益探索和尝试。本文将从非法行医专题分析入

〔收稿日期〕 2016-12-21

〔作者简介〕 王晖,高级工程师,发表论文30余篇。

手, 探讨卫生计生监督的大数据应用, 为降低非法行医案发率提供辅助决策依据。

2 卫生计生监督的数据需求与数据特征

2.1 数据需求

卫生计生监督信息报告系统, 采用 3 种方式将信息采集功能延伸到基层: 一是通过国家建设的卫生计生监督信息报告系统, 实现信息直报; 二是通过标准版业务系统, 借助手持终端, 实现现场信息填报; 三是与自建业务系统省份实现数据交换, 获取基层卫生计生监督业务数据。通过这 3 种方式, 可最大程度地提升被监督单位的监督覆盖面。一线卫生计生监督人员可以直接将结构化的信息提交到国家级卫生计生监督信息平台, 最大限度地提高信息采集的准确性与可靠性。近年来, 随着互联网应用的飞速发展, 电子商务、社交网络的广泛应用, 使得互联网中产生了人群活动的大量非结构化的公共卫生相关信息, 如何利用这些外部的、海量的、动态的信息, 是卫生计生监督大数据应用的新需求。

2.2 数据特征

卫生计生监督大数据的特征主要体现在数据量

大、业务逻辑复杂等方面。卫生计生监督涉及的专业包含公共场所、生活饮用水、职业卫生、放射卫生、学校卫生、医疗卫生、传染病防治、消毒产品、餐饮具消毒、血液安全、计划生育等。卫生计生监督的主要业务, 以被监督单位底档卡为基础, 实现日常性监督检查、监督抽检、案件查处等活动。卫生计生监督信息来源多样化, 包括信息在线填报、数据交换、设备终端的自动采集以及来自国家统计局年鉴的外部数据, 包括全国各省每年度的每万人卫生技术人员数、人口数、人口密度、医疗机构数、GDP 等。

3 卫生计生监督大数据应用架构

3.1 总体架构

卫生计生监督大数据总体架构, 从卫生计生监督的数据需求出发, 建立卫生计生监督大数据采集、存储、检索、计算、展现、管理的分布式基础数据平台。实现的业务包括即时统计分析、专题分析、建议与提醒和饮用水在线监测等。总体架构, 见图 1。

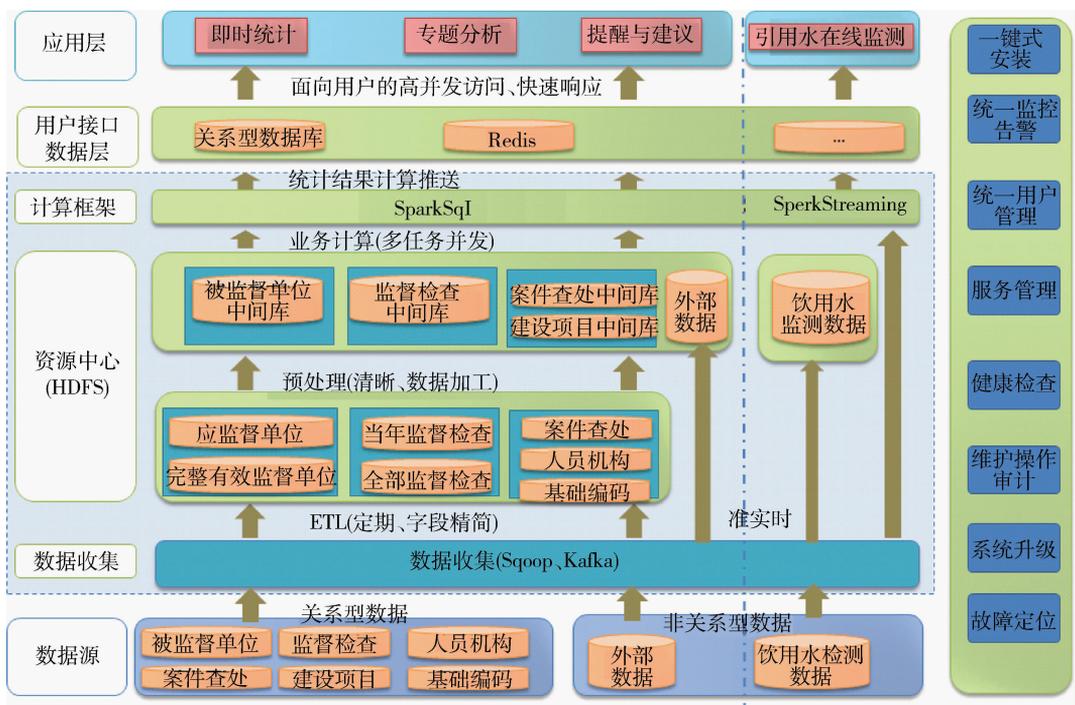


图 1 卫生计生监督大数据平台总体架构

3.2 技术架构

卫生计生监督大数据平台技术架构，见图 2。包含 4 部分：数据存储层、数据服务层、前端展现和分析挖掘工具箱。数据存储层主要包含以 HDFS 为代表的分布式文件系统和以 Oracle 为代表的关系型数据库两种存储方式。海量数据主要存储在分布式文件系统中，充分利用分布式文件系统易于扩展的特性；而汇总处理后的数据主要存储在关系型数据库中，以便于查找更新。数据服务层提供统一的数据访问接口。上层应用只需要遵循统一的数据访问规范即可获取底层数据，而无须关心数据的存储介质和位置。分析挖掘工具箱里放置 SparkR，可以方便卫生计生监督业务人员利用 R 语言直接访问数据存储中的数据，分析挖掘的结果，直接利用 SparkR 的展现功能进行展示。

构，见图 3，包括 4 台大数据节点外接 1 台 Oracle 关系数据库节点，各节点间采用两条千兆以太网互联，1 条千兆网络负责计算存储，1 条千兆网络负责监控管理。

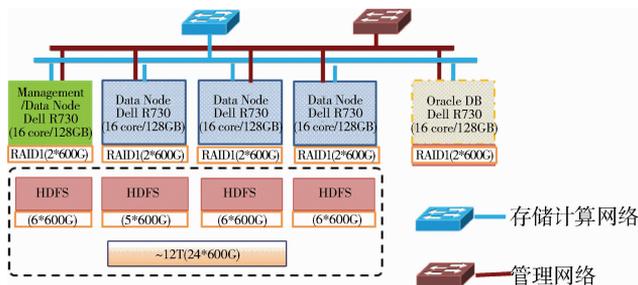


图 3 卫生计生监督大数据平台部署架构

4 非法行医大数据应用

4.1 应用目标

基于卫生计生监督大数据平台，通过多种采集方式获取外部数据，对采集的非法行医案件查处信息和外部接入的国家统计年鉴等数据进行相关动态分析与展示，分析出非法行医主要相关因素并为相关部门决策提供数据参考。

4.2 非法行医整体情况分析

对 2012 - 2015 年连续 4 年全国非法行医案件数进行分析，排名靠前的省份为山东、河南、安徽、上海、浙江、江苏、重庆、贵州、云南。从地域分布上来看，主要集中在华东、江浙、西南地区，并且有华东地区向西南地区扩散的趋势，见图 4。

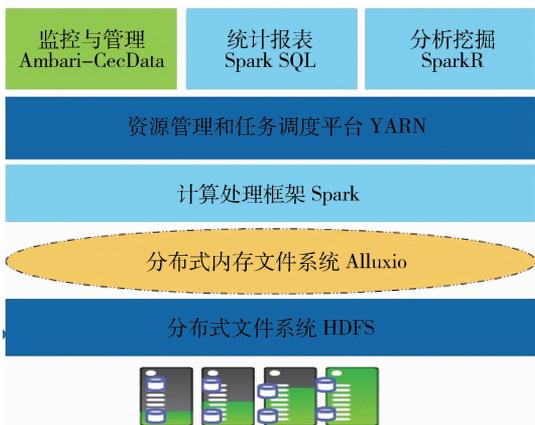


图 2 卫生计生监督大数据平台技术架构

3.3 部署架构

卫生计生监督大数据平台集群采用混合式架



图 4 非法行医地域分布图

4.3 非法行医相关因素分析

针对以上省份特点，首先从整体上对可能影响非法行医案件发生的人口数、人口密度以及每万人卫生技术人员数等相关因素进行分析，然后再对上述重点地区进行进一步的分类分析：华东地区（山东、安徽）：农村人口比例较大、当地人有特殊的就医习惯。江浙地区（浙江、上海）：经济较为发达，人口密度大、流动人口多。西南地区（云南、贵州）：医疗条件欠发达、少数民族地区，近年来地方执法打击力度较大。

4.3.1 与人口数的关系 对2012-2015年全国各省非法行医案件数进行分析，非法行医案件高发的山东、安徽、河南、浙江、江苏，都是全国人口数排名前10位的人口大省，见表1。通过对2014年各省非法行医案件数与各省人口数进行相关分析，可以看出大部分地区非法行医案件数与总人口数具有一定的正相关关系，见图5。

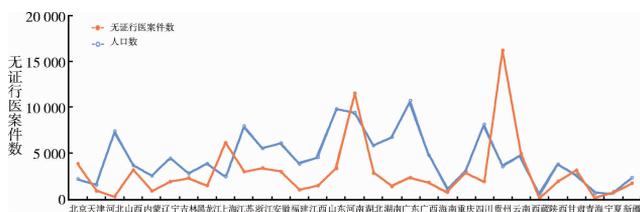


图5 2014年无证行医案件与人口数关系

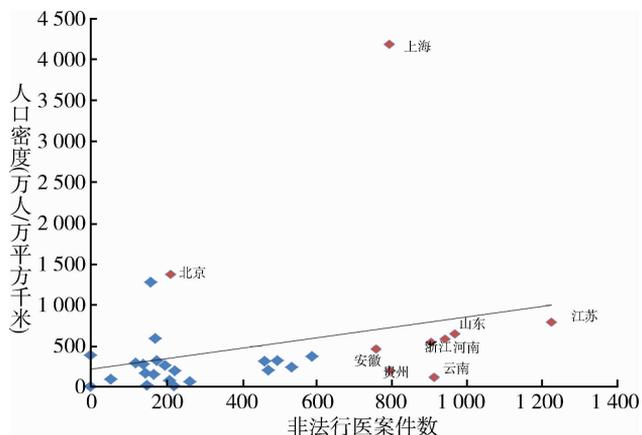
表1 2014年非法行医重点省人口全国排名

省份	人口数(万)	排名
山东	9 789	2
安徽	6 089	8
河南	9 436	3
浙江	5 508	10
江苏	7 960	5

注：人口数来源于《2015年国家统计局年鉴》。

4.3.2 与人口密度的关系 对2014年全国非法行医案件数与人口密度进行相关分析可以发现，大部分地区非法行医案件数与当地人口密度具有一定的线性相关性，案件发生率高的地区普遍人口密度大，见图6，如案件高发的上海、江苏、浙江、安徽、山东、河南等省份，人口密度排名都在全国的

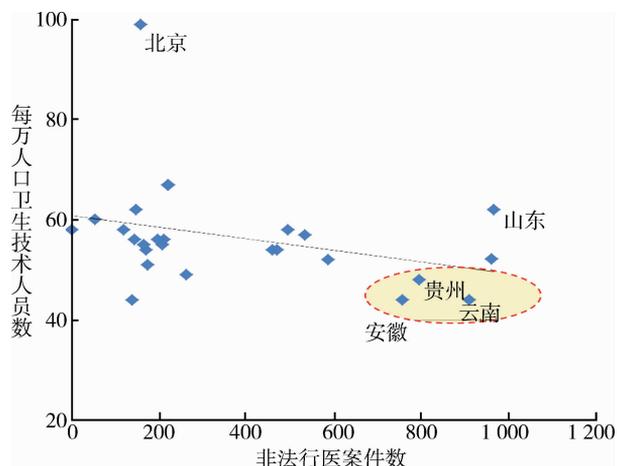
前几位。



注：人口密度基于《2015年国家统计局年鉴》
2014年各省人口总数和地域面积计算。

图6 非法行医与人口密度关系

4.3.3 与每万人口卫生技术人员数的关系 对2014年各省非法行医案件数与各省每万人口卫生技术人员数进行相关分析可以发现，大部分地区非法行医案件数与本省每万人口卫生技术人员配备数整体上有了一定的线性相关性，表现出每万人口卫生技术人员数配备较高的地区，非法行医案件数相对较少的趋势，见图7。



注：每万人口卫生技术人员数来源于
《2015年国家统计局年鉴》。

图7 非法行医与每万人口卫生技术人员数关系

4.3.4 特殊高发地区非法行医情况分析 通过进一步分析发现，华东地区、江浙地区、西南地区是重点高发地区，与人口密度、每万人口卫生技术人员

员等相关因素分析基本符合但还是有一定的特殊性，下面对以上地区再进行单独分析。(1) 华东地区非法行医与农村人口相关分析。通过对 2012 - 2015 年全国非法行医案件高发省份（安徽和山东）分析发现，安徽和山东非法行医案件涉案人员 90% 以上都为本地户籍人口。安徽非法行医案件主要集中在北部地区，山东非法行医主要集中在南部地区，见图 8，其中，安徽省特别是北部农村地区的案件占比超过 80%，山东以及南部地区农村地区案件占比也在 70% 左右。上述案件高发地区，均为人口多且农村人口占比高的地区，这些地区医疗资源配置普遍不高，并且当地居民有对非法行医危害性认识不足等问题^[1]。(2) 浙江、上海地区非法行医与流动人口因素分析。通过对 2012 - 2015 年浙江和上海两省涉案人员户籍分析，发现外省户籍占较高比例，且具有一定的地区聚集性。其中，浙江涉案人员中外省户籍占比均在 50% 以上，上海涉案人员外省户籍占比均在 90% 以上，且涉案人员的户籍主要分布在安徽、河南、江西等地，见图 9。这与浙

江、上海地区人口密度较大、经济活跃、外来务工人员较多等影响因素有一定的相关性^[2]，且外来务工人员更多选择非法行医点就诊。(3) 西南地区非法行医案件与医疗水平、打击力度的相关分析。通过对 2012 - 2015 年西南地区的非法行医案件数分析发现，近年来有明显上升的趋势。云南省和贵州省涉案人员主要为本地户籍，其中，云南省本地户籍占比在 60% ~ 70%，贵州省涉案人员本地户籍占比均在 80% 以上。根据对案件数、涉案人员户籍分布以及与国家统计局发布的各省每万人口卫生技术人员数相关分析，包括对当地地理环境和就医习惯等因素的分析，可知案件高发可能与以下因素有一定的相关性：①当地医疗资源相对不足。贵州、云南地区每万人口卫生技术人员数在全国来看，配比较低，医疗人员紧缺，并且当地多山交通不便，容易产生非法行医案件。②少数民族聚集地区，有着当地的就医习惯。③非法行医打击力度较大，云南 2013 年开始使用手持执法终端，贵州监督执法部门有严格执法任务。

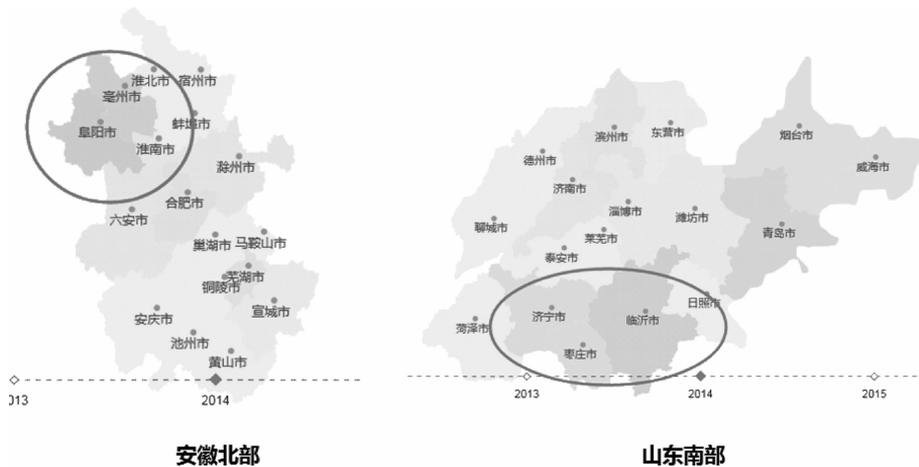


图 8 安徽、山东非法行医地区分布



图 9 全国各省涉案人员外省户籍分布

4.4 非法行医专题分析总结

通过对全国非法行医相关因素的分析，建议采取如下措施：(1) 加强农村、山区医疗卫生人力资源。(2) 加强流动人口的医疗保障措施。(3) 对不良就医行为进行宣传干预。(4) 对多省交界地区进行多地卫生计生监督机构联合执法。

(下转第 69 页)

- Nature Reviews Genetics, 2014, 15 (8): 556 – 570.
- 22 Frieze S, O'geen H, Littlepage LE, et al. Global Analysis of ZNF217 Chromatin occupancy in the Breast Cancer Cell Genome Reveals an Association with ERalpha [J]. BMC Genomics, 2014, (15): 520.
 - 23 National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease. Toward Precision Medicine: building a knowledge network for biomedical research and a new taxonomy of disease [M]. National Academies Press, 2012: 34 – 37.
 - 24 Katsnelson A. Momentum Grows to Make Personalized Medicine More Precise [J]. Nat Med, 2013, 19 (3): 249.
 - 25 Rowley J D. Letter: A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia Identified by Quinacrine Fluorescence and Giemsa Staining [J]. Nature, 1973, 243 (5405): 290 – 293.
 - 26 Ley TJ, Mardis ER, Ding L, et al. DNA Sequencing of a Cytogenetically Normal Acute Myeloid Leukaemia Genome [J]. Nature, 2008, 456 (7218): 66 – 72.
 - 27 International Cancer Genome C, Hudson TJ, Anderson W, et al. International Network of Cancer Genome Projects [J]. Nature, 2010, 464 (7291): 993 – 998.
 - 28 Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer [J]. Nucleic Acids Research, 2015, 43 (Database issue): D805 – D811.
 - 29 Bonham M D. BIREME: latin american and caribbean health sciences information center [J]. Bulletin of the Medical Library Association, 1990, 78 (2): 119 – 123.
 - 30 Ashworth A, Hudson TJ. Genomics: comparisons across cancers [J]. Nature, 2013, 502 (7471): 306 – 307.
 - 31 Huang T, Alvarez A, Hu B, et al. Noncoding RNAs in Cancer and Cancer Stem Cells [J]. Chinese Journal of Cancer, 2013, 32 (11): 582 – 593.
 - 32 Trepanier A, Ahrens M, Mckinnon W, et al. Genetic Cancer Risk Assessment and Counseling: recommendations of the national society of genetic counselors [J]. Journal of Genetic Counseling, 2004, 13 (2): 83 – 114.

(上接第 63 页)

5 结语

大数据给人的第一感觉是技术层面的事情，但大数据平台“鼓励从更多、更广、更深的角度看待、分析数据”的工作理念对工作的影响更为深远。大数据平台的应用促使人们从更广、更深的角度看待和分析数据，更加宏观地思考监督数据与监督之外的数据如何结合，如何通过更广泛的数据联系来发现可能的关系和数据价值，例如与国家宏观经济数据、卫生统计数据等之间进行关联分析，发现可能帮助开展卫生计生监督工作的有意义的结论。大数

据平台在大数据管理和利用方面提供了一个扩展性、可管理性、集成性方面都比较好的技术手段。大数据平台可以让业务人员更方便去思考 and 实现数据的多角度利用，便于从概念验证转向生产系统。

参考文献

- 1 樊春, 徐庆华, 胡庆生, 等. 安徽省 3 市黑诊所周边居民对非法行医认知及其影响因素研究 [J]. 中国卫生监督杂志, 2015, 22 (3): 262 – 266.
- 2 涂惠敏. 浙江省非法行医现状调查及长效治理策略分析研究 [D]. 杭州: 浙江大学, 2008.