

## ·新型冠状病毒肺炎疫情防控·

# 宁波市鄞州区健康大数据平台在新型冠状病毒肺炎疫情病例发现中的应用探索

孙烨祥<sup>1</sup> 吕筠<sup>2</sup> 沈鹏<sup>1</sup> 张敬谊<sup>3</sup> 路平<sup>3</sup> 黄文贊<sup>1</sup> 林鸿波<sup>1</sup> 水黎明<sup>4</sup> 李立明<sup>2</sup>

<sup>1</sup>宁波市鄞州区疾病预防控制中心数据中心 315100; <sup>2</sup>北京大学公共卫生学院流行病与卫生统计学系 100191; <sup>3</sup>万达信息股份有限公司,上海 200000; <sup>4</sup>宁波市鄞州区卫生健康局 315100  
孙烨祥和吕筠对本文有同等贡献

通信作者:水黎明, Email: 70776165@qq.com; 李立明, Email: lmlee@bjmu.edu.cn

**【摘要】** 在新型冠状病毒肺炎(COVID-19)疫情防控中,在没有疫苗和特异性治疗药物的情况下,控制传染源成为遏制疫情流行的最重要的防控措施之一。宁波市鄞州区在积极落实传统“早发现”综合措施的同时,通过联防联控机制实现了部门间数据信息共享,依托融合了医疗、疾控以及非卫生部门数据的健康大数据平台,创新性地探索开展大数据驱动的线上可疑病例筛选、线下核实处置的COVID-19病例发现工作模式,为今后实现更有效和高效的动态、持续的传染病监测预警奠定工作基础、积累经验。本研究对宁波市鄞州区的这一模式探索进行介绍,并对大数据驱动的监测模式在传染病防控中的作用进行讨论。

**【关键词】** 新型冠状病毒肺炎; 大数据; 病例发现; 范式

**基金项目:**国家自然科学基金(91846303); 北大百度基金——面向人群健康和重大疾病的大数据平台建设研究(2019BD010); 宁波市鄞州区科技局科技计划(2019-63-34)

DOI:10.3760/cma.j.cn112338-20200608-00818

## Application of healthcare big data in active case finding of COVID-19 in Yinzhou district of Ningbo

Sun Yexiang<sup>1</sup>, Lyu Jun<sup>2</sup>, Shen Peng<sup>1</sup>, Zhang Jingyi<sup>3</sup>, Lu Ping<sup>3</sup>, Huang Wenzan<sup>1</sup>, Lin Hongbo<sup>1</sup>, Shui Liming<sup>4</sup>, Li Liming<sup>2</sup>

<sup>1</sup>Department of Data Center, Yinzhou District Center for Disease Control and Prevention, Ningbo 315100, China; <sup>2</sup>Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China; <sup>3</sup>Wonders Information Coompany Limited, Shanghai 200000, China; <sup>4</sup>Yinzhou District Health Bureau, Ningbo 315100, China

Sun Yexiang and Lyu Jun contributed equally to the article

Corresponding authors: Shui Liming, Email: 70776165@qq.com; Li Liming, Email: lmlee@bjmu.edu.cn

**【Abstract】** During the prevention and control of the COVID-19 epidemic, identifying and controlling the source of infection has become one of the most important prevention and control measures to curb the epidemic in the absence of vaccines and specific therapeutic drugs. While actively taking traditional and comprehensive “early detection” measures, Yinzhou district implemented inter-departmental data sharing through the joint prevention and control mechanism. Relying on a healthcare big data platform that integrates the data from medical, disease control and non-health sectors, Yinzhou district innovatively explored the big data-driven COVID-19 case finding pattern with online suspected case screening and offline verification and disposal. Such effort has laid a solid foundation and gathered experience to conduct the dynamic and continuous surveillance and early warning for infectious disease outbreaks more effectively and efficiently in the future. This article introduces the exploration of this pattern in Yinzhou district and discusses the role of big data-driven disease surveillance in the prevention and control of infectious diseases.

**【Key words】** COVID-19; Big data; Active case finding; Paradigm

**Fund programs:** National Natural Science Foundation of China (91846303); Research on the Construction of Big Data Platform for People's Health and Major Diseases Supported by Baidu Foundation of Peking University (2019BD010); Science and Technology Program of Yinzhou District Science and Technology Bureau of Ningbo City (2019-63-34)

DOI:10.3760/cma.j.cn112338-20200608-00818

新型冠状病毒肺炎(COVID-19)疫情是新中国成立以来我国经历的传播速度最快、感染范围最广、防控难度最大的一次重大突发公共卫生事件。在没有疫苗和特异性治疗药物的情况下,控制传染源成为遏制疫情最重要的防控措施之一。《新型冠状病毒肺炎防控方案》中明确要求各地积极采取措施,做到“早发现、早报告、早隔离、早治疗”,控制疫情传播<sup>[1]</sup>。

宁波市鄞州区区域面积 814.2 km<sup>2</sup>,常住人口 108 万,下辖 12 个街道、9 个镇。近 5 年来,鄞州区卫生健康局全力打造区级全民健康信息平台和健康大数据平台,研发了基于大数据的分层分级分域数据管理技术体系、基于敏感信息规避和触发技术的数据隐私保护技术等数据治理技术,采用可配置的数据质量修复融合技术、以患者为中心的全链条疾病特征自动识别等技术,具有大数据构架安全便捷、多源异构数据的标准化融合能力。两大平台同时授权给鄞州区 CDC 开展健康大数据驱动下的公共卫生实践与疾病管理模式探索。

在此次抗击 COVID-19 疫情期间,鄞州区在积极落实传统“早发现”综合措施的同时,通过联防联控机制实现了部门间数据信息共享,依托融合了医疗、疾控以及非卫生部门数据的健康大数据平台,开发了“新型冠状病毒肺炎监测与预警信息系统”,创新性地探索开展大数据驱动的线上可疑病例筛选预警、线下核实处置的 COVID-19 病例发现工作模式,为今后实现更有效和高效的动态、持续的传染病监测预警奠定工作基础、积累经验。

本研究将对鄞州区健康大数据平台中疫情相关的数据来源和内容、疫情防控中病例发现的主要途径及大数据应用平台在其中的应用进行简单介绍,然后对这种大数据驱动的监测模式在传染病防控中可能发挥的作用、有待完善之处及未来的应用发展进行讨论。

### 一、健康大数据平台的数据来源和内容

鄞州区通过联防联控机制,实现了卫生和非卫生多个部门、众多机构的数据共享,将所有数据导入鄞州区健康大数据平台,大部分数据可以实现每日更新。万达信息股份有限公司大数据团队免费开发模型代码嵌入健康大数据平台,进行数据的多源异构融合和 COVID-19 监测模型的自动运算。

1. 宁波市市级医疗机构、鄞州区 532 家区、街道两级医疗机构(包括综合性公立医院、专科医院、妇幼保健院、社区卫生服务中心、社区卫生服务站等)和部分民营健康管理服务机构的数据。

(1)居民健康档案:鄞州区 102.4 万居民建立了健康档案,建档率达 95%,可掌握辖区绝大多数居民的基本个人信息。

(2)临床诊疗数据:区内建档人群在市、区、街道三级医疗机构就诊时的电子病历数据,包括门诊电子病历、住院病案首页、检验检查数据、影像学数据、药物使用数据等;区内未建档人群在区、街道两级医疗机构就诊时的门诊电子病历数据。区、街道两级医疗机构的法定传染病报告数据。

(3)送检样本的检测数据和隔离人员的健康监测数据:第三方检测机构收取区内医疗机构送检样本的病毒核酸和抗体检测结果,可推送给医疗机构的实验室信息系统(LIS);疫情期间作为集中隔离点的健康管理中心、酒店等机构针对感染者、密切接触者或重点人群的健康监测数据。

### 2. 疾控机构的传染病报告和流行病学调查数据:

(1)传染病报告数据:鄞州区居民被全国各级医疗机构诊断为 COVID-19 病例,数据均可通过国家传染病与突发公共卫生事件监测信息系统(网络直报系统)推送至鄞州区 CDC,新诊断的病例按日导入大数据平台。

(2)流行病学调查数据:鄞州区 CDC 针对 COVID-19 病例开展的流行病学调查数据当日导入大数据平台;国家、省、市各级 CDC 针对鄞州区居民开展的密切接触者流行病学调查数据,鄞州区 CDC 在数据接收当日导入大数据平台。

3. 在浙江省启动重大突发公共卫生事件一级响应状态下,由鄞州区 COVID-19 疫情联防联控办公室协调共享的重点人群名单:

(1)公安部门:根据出行记录核查发现的有高风险地区驻留史的鄞州区户籍人员以及 2019 年曾居住在鄞州区的高风险地区户籍的流动人员。公安部门共享每日更新的人员名单和基本个人信息。

(2)交通管理部门:从高风险地区经公共交通抵达宁波市的乘客名单。更新数据每日与卫生部门共享,导入大数据应用平台。

(3)教育部门:由教育部门提供区内学校(初中、小学、幼托机构)学生及教职工名单,并提供学校人员信息维护客户端和学生及教职工健康信息查询功能,方便数据的动态更新,开展在读学生群体的传染病持续监测。此外,共享前往高风险地区就读大学的宁波籍学生名单,用于特定学生群体的相关可疑病例监测。

(4)统计部门:共享鄞州区各街道/镇人口数据,

可用于区域性指标的计算。

4. 从浙江省大数据发展管理局获取的“健康码”数据：与全省“健康码”管理信息系统对接，每日更新“健康码”数据，可识别学生所在学校和班级，用于复学后对辖区内在读学生的健康监测。

## 二、疫情防控中 COVID-19 病例发现的主要途径

抗击疫情期间，鄞州区积极落实传统的病例“早发现”的综合措施。如图 1 中右侧显示，病例可通过多种途径被发现。个体因病就诊：由医疗机构发现并诊断；疫情期间医疗机构会对发热患者进行采样检测。口岸检疫措施：入境人员主动申报自己有相关的临床症状、在高风险地区的驻留史或可能的接触史，或经体温检测、医学巡查发现，进一步医学排查判断为可疑人员的转送指定医院进行诊疗。隔离人员健康监测：对按规定进行居家或集中隔离的人员，由医务人员进行健康监测，对出现相关临床症状者进行医学排查，判断为可疑人员的转送指定医院进行诊疗。社区入户排查：组织基层工作人员对社区居民进行入户排查登记，对发热人员或有流行病学史人员进行相应处置。个体自主申报可疑线索：宁波市卫生健康局开发了针对居民的“新型冠状肺炎可疑线索自报系统”，用于个人关于可能的流行病学史及健康情况的申报，也是一种病例发现途径。

在上述综合病例发现途径的基础上，鄞州区又探索开展了大数据驱动的可疑病例筛选（图 1 左侧显示）。凭借联防联控、信息共享的优势，区公安、交通管理、教育等非卫生部门与卫生部门共享有高风险地区旅行史或驻留史的人员名单，或存在高风险流行病学史（即旅行史、驻留史或与高风险地区人员的接触史）的可能性更高的人员名单（如在高风险地区就读的宁波籍大学生）；疾控机构有 COVID-19 病

例的密切接触者名单。上述重点人群以外的社区居民可通过居民健康档案获取名单。由于就医，以及航空、铁路购票均已采取实名制管理，绝大多数人群可通过身份证号码匹配医疗机构的临床诊疗数据，筛选可疑病例。如果发现的可疑病例为非鄞州区居民，将相关数据推送给属地化疾控机构。

通过文献法和德尔菲法，依据国家卫生健康委员会在 1 月 22 日印发的《新型冠状病毒感染的肺炎诊疗指南（试行第三版）》，制定了可疑病例的筛选条件：① ICD-10 疾病编码定义的“肺炎”“病毒性肺炎”等 16 类疾病类别，除外“新型冠状病毒感染的肺炎”；② 有发热、咳嗽、乏力、气促等临床表现之一；③ 血常规提示淋巴细胞计数降低，或白细胞计数正常或降低；④ 具有肺部感染等肺炎影像学表现。由于重点人群和普通社区居民存在高风险流行病学史的概率不同，故设定了不同的筛选定义：重点人群：满足上述筛选条件中①或②的任何一条。社区居民：满足上述筛选条件中①或②的任何一条，且同时满足③和④；即需同时符合临床、实验室和影像学表现。可见，重点人群采用了敏感性更高的筛选定义；而针对社区居民的筛选定义则考虑提高特异性，控制假阳性。

对不同类型人群按相关规定设置监测期限：① 相关日期后 2 周：密切接触者、高风险旅行史、因肺炎相关疾病住院治疗后出院者（COVID-19 检测阴性）；② 相关日期后 3 周：有高风险地区驻留史的鄞州区户籍人员；③ 持续监测：在高风险地区就读的宁波籍大学生、2019 年曾居住在鄞州区的高风险地区户籍的流动人员、普通社区居民。重点人群在上述对应的监测期限内，采用重点人群的筛选定义；在监测期限外，则按社区居民的筛选定义持续监测。

### 大数据应用平台每日 0 点运行可疑病例的筛选

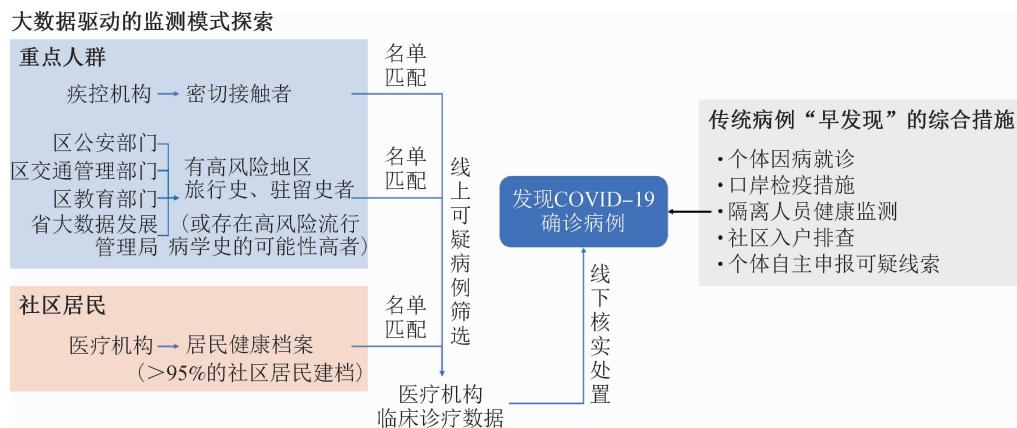


图 1 疫情防控中 COVID-19 病例发现的主要途径

模型。经筛选符合可疑病例定义的人员名单被推送至“新型冠状病毒肺炎监测与预警信息系统”，属地化医疗机构的公共卫生人员可通过专属的疾病监测系统网页浏览推送的人员基本个人信息，以及医疗机构就诊、检查检验、临床诊断等信息。公共卫生人员需当日对可疑病例逐个电话或入户快速核实其流行病学史、自出现症状以来的病情变化情况（如呼吸道症状有无缓解、体温变化等）、是否接受过市级医疗卫生机构的核酸检测及其结果（区内医疗卫生机构的核酸检测结果可通过数据平台匹配），并及时将核实信息补充到该信息系统的可疑病例个案记录中。疾控部门联合医疗机构负责对可疑病例进一步接受核酸检测的必要性进行研判和实施检测。同一个人可能多次到医疗机构就诊，还可能是到不同医疗机构就诊，只要满足可疑病例的定义，系统就会推送个案预警；每次推送，公共卫生人员都需要进行线下核实，并了解疾病发展变化情况。

截至2020年3月30日，鄞州区共报告13例COVID-19确诊病例。其中，5例由大数据应用平台根据前一日医疗机构就诊信息筛选为可疑病例；2例为密切接触者，1例为高风险地区就读的宁波籍学生，2例为普通社区居民。经线下核实，4例在就诊当天已按接诊医生处置接受了核酸检测，另1例在就诊的次日接受了核酸检测，随后根据检测结果确诊为COVID-19病例。另有几例未被大数据平台筛选为确诊病例，或因呼吸道以外问题就诊，或被发现为密切接触者或聚集性疫情相关的密切接触者，或为轻症、无症状感染者等，接受核酸检测后被确诊。

### 三、大数据驱动的监测模式在传染病防控中的作用

COVID-19作为一种新发传染病，对其病原学和流行病学特征、临床表现等认识还远远不够。所以在大数据应用平台中设定的可疑病例的筛选定义仍存在不足之处。例如，目前的临床标准仍以呼吸道疾病表现为主，同时结合血常规和影像学表现，而鄞州区确诊的半数以上病例的血常规和影像学表现不明显，少数病例因呼吸道以外疾病就医。鄞州区在疫情期间累计报告确诊病例13例，数量不多，属于低流行地区；筛选发现的可疑病例中，最终确诊病例的比例并不高，即阳性预测值低。另外，此次疫情受到全社会的高度重视，病例发现的各种措施得以加强。有些病例在尚未出现明显临床症状并到医疗机构就诊时，就已通过密切接触者调查被发现。而

健康大数据平台的核心数据是临床诊疗数据，其他来源的数据主要用来反映流行病学暴露史。个体未到医疗机构就诊的情况下是无法通过这种方式被发现的。总的来说，在疫情防控应急状态的综合防控措施下，这种大数据驱动的监测模式未成为低流行地区发现病例的主要途径；但是，它为我国现代化传染病监测体系的发展探索和示范了一条新途径。

传染病病例的发现通常根据流行病学史和临床表现综合分析，进一步根据病原学或血清学证据进行诊断。传统工作模式普遍存在以下问题，在非疫情时代尤为突出：个体就医时临床表现不典型，对相关的流行病学史不清楚或谎报瞒报；接诊医生因为意识或能力问题，未能及时发现、诊断、报告可疑病例；个体连续多次就诊，由于接诊医生不同，或就诊于不同科室、不同医院，难以结合之前的诊治情况和病情发展进行综合分析；口岸检疫措施适用于境外人员入境当时，可能尚未出现可观察到的临床症状或故意掩盖症状，对相关的流行病学史不清楚或谎报瞒报。

鄞州区基于健康大数据平台的这种工作模式的优势主要表现：①利用区域范围内临床诊疗数据进行事后可疑病例筛选，构建了落实医疗机构病例早发现的第二道防线，防范各级各类医疗机构因管理、培训不到位以及医务人员诊断、报告意识和能力的参差不齐而导致的疏漏。②跨机构、跨区域的健康医疗数据融合，有机会根据患者短期内连续就医的信息进行综合分析判断。③旅行、居住史通常可以反映重要的流行病学史信息，通过公安、交通管理、教育等跨部门、跨行业的数据共享融合，可以部分地解决相关信息自报存在的谎报瞒报情况，也可用于快速搜索密切接触者。④在大数据应用平台上，根据客观记录的旅行、居住史可区分具有不同暴露风险的人群，有效标记重点人群，设置敏感性和特异性不同的临床筛选标准。随着对疾病认识的更新以及疫情防控进展、聚集性疫情和当地社区传播风险等形势的发展变化，也可及时在平台中修改筛选标准。⑤利用各部门现有的数据信息，有效地将基层人员从社区排查时逐一询问、手动填表、重复填表的繁重工作任务中解放出来，也克服了基层数据采集中存在的数据误报瞒报、质量参差不齐、时效性低等问题。目前，根据旅行、居住史信息确定的重点人群名单可实现每日更新，大大提高了病例发现的工作效率。

当然，对这种大数据驱动的传染病监测模式的

探索还远远不够,仍有很大的完善空间。下一步,可加强这种模式在更多法定报告传染病监测中的应用探索。对于可疑病例的筛选标准,需综合权衡发现可疑病例的能力和线下核实的工作量,不断优化筛选模型的敏感性和特异性。此次 COVID-19 疫情中,关于可疑病例的线上筛选仍主要基于单次就诊时的诊疗数据;对于短期多次就诊者,由属地化公共卫生人员调阅医疗记录进行综合判断。今后,系统可继续优化,对这类多次就诊的人员进行自动标记和研判。目前,可疑病例的筛选主要基于诊疗数据中的临床诊断、症状、实验室检查和影像学诊断结果;今后还可以探索用药数据、电子病历中特异性症状描述等更多类型数据的可用性;也可增加鉴别诊断功能,先排除可明确诊断的其他疾病。另外,可以考虑利用区域范围的大数据平台,识别短期内多次就诊者,特别是在不同医疗机构就诊者,识别有特定流行病学史者,整合到医院信息管理系统中,利用“白名单”技术推送给医疗机构接诊医生,做必要的提醒预警。本次疫情中,未能直接获得归属上级教育部门主管的高中学校学生和教职工数据,提示跨层级的联防联控机制仍有待加强。非卫生部门共享的是特定重点人群的名单和基本个人信息;希望今后可以探索安全有效的方式实现更多数据的共享,促进流行病学调查手段的现代化发展,提升疾病的人群溯源和密切接触者的追踪能力,以及集体单位、居住区、公共场所活动区内暴发疫情的识别预警。

#### 四、小结

鄞州区经过多年的努力,在跨机构和跨区域(即

市区间)健康医疗大数据的共享、汇聚、融合,以及健康大数据平台的建设发展上已经初见成效。此次 COVID-19 疫情中,更是借助联防联控的工作机制实现了跨部门数据的共享。在大数据驱动的可疑病例筛选预警方面进行了创新性的探索,为我国传染病监测现代化发展示范了一条新的技术路径。另外,鄞州区正在计划基于大数据平台,探索将疾病监测预警模块数据直接推送至医疗机构的路径,利用区域数据辅助临床诊疗与可疑病例的排查监测,使基于大数据的疾病监测与医疗实践之间形成闭环,加强医防协同。

我们也希望这种政府层面的、针对公共卫生工作的多部门协调工作机制平台在后疫情时代仍可以常态化推进。政府部门间的数据共享不会因为疫情结束而终止,此次疫情应该成为打通“数据孤岛”的破冰之旅。

**利益冲突** 所有作者均声明不存在利益冲突

#### 参 考 文 献

- [1] 国家卫生健康委员会疾控局.新型冠状病毒肺炎防控方案(第六版)[EB/OL].(2020-03-07)[2020-05-30].<http://www.nhc.gov.cn/jkj/s3577/202003/4856d5b0458141fa9f376853224d41d7.shtml>. Bureau of Disease Control and Prevention, National Health Commission. Protocol on Prevention and Control of Novel Coronavirus Pneumonia (Edition 6) [EB/OL]. (2020-03-07) [2020-05-30]. <http://www.nhc.gov.cn/jkj/s3577/202003/4856d5b0458141fa9f376853224d41d7.shtml>.

(收稿日期:2020-06-08)

(本文编辑:万玉立)

#### · 征订启事 ·

## 本刊 2021 年征订启事

《中华流行病学杂志》创刊于 1981 年,是由中华医学会主办、中国疾病预防控制中心传染病预防控制所承办的流行病学及其相关学科的专业学术期刊,以从事预防医学与公共卫生、基础医学、临床医学的广大工作者为读者对象。报道内容和栏目设置涵盖流行病学及其各分支学科的科研成果、疾病预防控制、大型队列研究、现场流行病学调查和监测、临床流行病学、分子流行病学、循证和转化医学、健康大数据、实验室研究、基础理论与方法、系列讲座、系统综述、经典案例、教育教学方法与实践等。

《中华流行病学杂志》被 Medline/PubMed、Scopus、CA、Europe PMC、中国科技核心期刊(中国科技论文统计源期刊)、中文核心期刊要目总览(北大核心目录)、中国科学引文数据库(CSCD)等多种国内外知名的检索系统和数据库收录。《中华流行病学杂志》近年连续被评为“百种中国杰出学术期刊”、“中国最具国际影响力学术期刊”、“中国国际影响力优秀学术期刊”、“中国精品科技期刊”、“RCCSE 中国权威学术期刊”等。有多篇论文入选“中国百篇最具影响国内学术论文”、“中国科协科技优秀论文”、“百篇中华医学优秀论文”和“F5000 论文”等。2019 年入选“中国科技期刊卓越行动计划”。

《中华流行病学杂志》全年出版 12 期,每期定价 35 元,全年 420 元,由全国各地邮局统一订阅,邮发代号:2-73;可登录中华医学网(<http://medline.org.cn/>)的“商城”(<http://medline.org.cn/mall/index.do>)和微信公众号“中华医学会杂志社员俱乐部(微信号:cmaclub)”的“商城”进行订阅。中华流行病学杂志编辑部地址:北京昌平区昌百路 155 号传染病所 B115 室,邮编:102206,电话(传真):010-58900730,Email:[zhlxz1981@sina.com](mailto:zhlxz1981@sina.com)。欢迎广大读者踊跃投稿(<http://chinaepi.icdc.cn>),积极订阅。