

# 基于 VOSviewer 软件的信息流行病学领域研究热点与主题分析

刘艳丽<sup>1</sup> 颜凡翔<sup>2</sup> 吴燕萍<sup>2</sup>

<sup>1</sup>杭州医学院图书馆, 杭州 311399; <sup>2</sup>杭州医学院公共卫生学院 (食品科学与工程学院), 杭州 311399

通信作者: 吴燕萍, Email: w\_yp@zju.edu.cn

**【摘要】** **目的** 了解全球信息流行病学相关研究主题、热点和发展趋势。**方法** 检索截止到 2023 年 3 月 3 日 Web of Science 核心合集数据库中收录的信息流行病学相关文献。利用 VOSviewer 1.6.19 可视化分析工具总结纳入文献的年度发表量、出版期刊、研究机构及作者信息等;并对出现频次前 100 的关键词进行共现聚类分析,分析研究热点主题与趋势。**结果** 截止到 2023 年 3 月 3 日,共检索到 14 521 篇文献。文献年度发文量 2001 年超 100 篇且呈逐年平稳增加趋势,2020 年发表文献量由于 COVID-19 疫情出现了成倍增长且 2021 年达到峰值 (2 593 篇),2022 年有所下降 (1 973 篇);出版文献数量前 10 位的期刊来自 5 个国家,出版量总计 2 099 篇 (14.46%, 2 099/14 521);发表文献数量前 10 位的研究机构 (除 WHO 外) 来自 4 个国家,发文量总计 2 943 篇 (20.27%, 2 943/14 521);发表文献数量前 10 位的作者来自中国和英国,发文量均在 40 篇以上,总计 566 篇 (3.90%, 566/14 521);高频关键词聚类于 4 个重要领域[艾滋病(HIV/AIDS)、流行病学(epidemiology)、流行病(epidemic)和 COVID-19]和 3 个热点主题(艾滋病公共卫生监测、流行病信息监测和 COVID-19 信息流行病)。**结论** 当前信息流行病学的研究热点集中于 COVID-19 的信息流行病管理,包括 SARS-CoV-2、信息流行病、精神卫生和社交媒体等内容。今后将围绕公共卫生焦点问题以信息流行病管理方式开展流行病学相关研究。

**【关键词】** 医学信息学;信息流行病学;文献计量学;研究热点;聚类分析

**基金项目:** 浙江省科技厅软科学研究计划 (2022C35042);浙江省医药卫生科技计划青年人才计划 (2019RC146);浙江省教育厅 2022 年省级课程思政示范课程《生物信息》;杭州医学院 2022 年省级大学生创新创业训练计划(S202213023048)

DOI:10.3760/cma.j.cn331340-20221214-00265

## Research hotspots and thematic analysis of infodemiology based on VOSviewer software

Liu Yanli<sup>1</sup>, Yan Fanxiang<sup>2</sup>, Wu Yanping<sup>2</sup>

<sup>1</sup>Library of Hangzhou Medical College, Hangzhou 311399, China; <sup>2</sup>School of Public Health, Hangzhou Medical College, Hangzhou 311399, China

Corresponding author: Wu Yanping, Email: w\_yp@zju.edu.cn

**【Abstract】** **Objective** To understand the topics, hotspots and trends of researches in the field of infodemiology. **Methods** Based on Web of Science Core Collection, the literatures related to infodemiology were collected until March 3rd, 2023. With VOSviewer version 1.6.19, published features including numbers of annual published articles, journal, organization and author information were analyzed. Based on the top 100 keywords in frequency, cluster analysis was used to analyze the research hotspots and trends. **Results** Until March 3rd, 2023, a total of 14 521 articles were retrieved. The annual number of published articles exceeded 100 in 2001 and showed an increasing trend in following years. In 2020, the number increased doubly due to the COVID-19, and reached the peak in 2021 (2 593 articles) and declined in 2022 (1 973 articles). The top 10 journals in publication number came from 5 countries, with a total of 2 099 articles (14.46%, 2 099/14 521). The top 10 research institutions (excluding World Health Organization) came from 4 countries, with a total of 2 943 articles (20.27%, 2 943/14 521). The top 10 authors publishing over 40 articles per person were from China and Britain, with a total of 566 articles (3.90%, 566/14 521).

According to high frequency keywords, 4 key research fields including HIV/AIDS, epidemiology, epidemic and COVID-19, and the 3 hotspots including HIV/AIDS public health monitoring, epidemiology monitoring and COVID-19 information epidemic were clustered. **Conclusions** The current hotspot of infodemiology focuses on the infodemic management of COVID-19 including SARS-CoV-2, infodemiology, mental health and social media. In the future, the research will focus on the public health and carry out epidemiological research in the way of information epidemic management.

**[Key words]** Medical informatics; Infodemiology; Bibliometrics; Research hotspots; Cluster analysis

**Fund Program:** Soft Science Research Program of Department of Science and Technology of Zhejiang (2022C35042); Young Talent Program of Zhejiang Medical Science and Technology Project (2019RC146); Curriculum Ideological and Political Demonstration Course 'Bioinformatics' supported by Department of Education of Zhejiang Province in 2022; Students' Innovation and Entrepreneurship Training Program in Zhejiang Province in 2022 supported by Hangzhou Medical College (S202213023048)

DOI:10.3760/cma.j.cn331340-20221214-00265

信息流行病学 (infodemiology) 是伴随着互联网的发展而衍生出的一门新兴交叉学科, 最初由信息 (information) 和流行病学 (epidemiology) 两个单词合成而来<sup>[1]</sup>, 与之相关的合成词还包括信息流行病 (infodemic) 和信息监测 (infoveillance)<sup>[2]</sup>, 后逐步发展为信息流行病学 (infodemiology)<sup>[3]</sup> 和信息流行病 (infodemic)<sup>[4]</sup>, 主要经历了三个发展阶段: 萌芽期 (1996—2005 年), 重建期 (2006—2019 年) 和形成期 (2020 年至今)。2020 年 6 月 30 日 WHO 将信息流行病学定义为管理信息流行病的科学<sup>[5]</sup>, 标志着其作为一门学科被全球卫生领域认可<sup>[6]</sup>。COVID-19 疫情暴发使全球对信息流行病 (或称信息疫情) 的关注呈现出井喷式发展, 但目前国内外对信息流行病学这一学科的研究仍有限<sup>[7-10]</sup>。本文对信息流行病学的概念进行了溯源, 并通过可视化软件分析识别出本领域的重要主题与研究热点, 分析其未来研究的发展方向。

## 资料与方法

### 一、数据来源

以 Web of Science 核心合集作为数据来源进行检索, 检索式为 TS=(infodemi\* OR infoveillance OR disinfectant) OR ((information OR misinformation OR disinformation OR malinformation) AND epidemic\*), 时间范围截至 2023 年 3 月 3 日。文献纳入标准: (1) 与信息流行病、信息流行病学、信息监测、虚假

信息、流行病学等主题相关; (2) 语言为英文; (3) 文献类型为论文与综述。排除标准: (1) 与本主题不符、内容无关文献; (2) 重复文献; (3) 报告、会议通知和约稿信息、在线发表文献、作者信息和单位信息未知文献。最终纳入相关文献 14 521 篇。

### 二、研究工具

VOSviewer 软件是基于 Java 语言开发的一款文献计量的可视化软件, 支持多种文献数据库, 支持作者共现分析、关键词共现分析、共被引文献 (来源期刊和作者) 分析等。节点类型可选择为发表时间、出版刊物、研究机构和文献等。其中, 节点大小代表关键词出现频次, 节点越大, 频次越高; 两个节点间的连线代表关键词之间的共现关系, 连线越粗, 二者间的关系越紧密, 节点越靠近, 则二者的共现频次越高。

### 三、研究方法

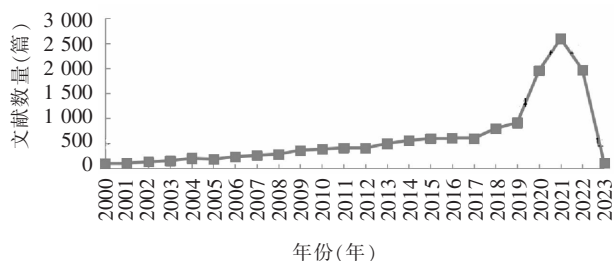
将 Web of Science 核心合集数据库中检索到的文献以纯文本格式导入到 VOSviewer 1.6.19 软件, 统计纳入文献的年度发表量、出版期刊、研究机构及作者信息等; 对出现频次前 100 的关键词进行共现的聚类分析, 得到信息流行病学关键词的共现网络视图和标签视图。

## 结 果

### 一、年度发表数量趋势分析

纳入的 14 521 篇信息流行病学文献年度发表量逐年增加, 2001 年开始年度发表量均超过 100

篇,且呈逐年平稳增加趋势;随着 COVID-19 疫情的暴发,2020 年发表文献量出现了翻倍(1 954 篇),2021 年文献数量平稳增加达到峰值(2 593 篇),2022 年文献数量相较于 2021 年有所下降(1 973 篇);2023 年截至 3 月 3 日,发文量为 112 篇(见图 1)。



注:数据截至 2023 年 3 月 3 日

图 1 信息流行病学研究文献发表数量的年度趋势

## 二、纳入研究文献的期刊分析

纳入的 14 521 篇文献来自 3 842 种期刊。其

中,发文量前 10 位的期刊主要来自美国(2 种)、加拿大(2 种)、瑞士(2 种)、英国(3 种)和荷兰(1 种);前 10 位期刊以中科院医学、信息及综合学科分区的 2 区和 3 区期刊为主,医学类(7 种)、信息工程类(1 种)及综合学科期刊(2 种)出版文献占比分别为 9.11%、0.78%和 4.58%;前 10 位期刊的发文量均在 100 篇以上,总计 2 099 篇(14.46%, 2 099/14 521)(表 1)。

## 三、纳入研究文献的研究机构

纳入的 14 521 篇文献中,共有 13 536 家机构参与信息流行病学研究。其中,发文量前 10 位的研究机构来自美国(4 家)、英国(2 家)、法国(1 家)和中国(2 家),这些机构的累积发文量除 WHO 外,分别为 1 307 篇(美国,9.00%, 1 307/14 521)、631 篇(英国,4.35%, 631/14 521)、369 篇(法国,2.54%, 369/14 521)和 368 篇(中国,2.53%, 368/14 521),前 10 位的机构共发

表 1 信息流行病学相关文献发表量排名前 10 位的期刊

序号	刊名	出版国家	医学(信息)分区 <sup>a</sup>	纳入文献量(篇) <sup>b</sup>
1	PLoS One	美国	综合 3 区	489
2	Journal of Medical Internet Research	加拿大	医学 2 区	290
3	International Journal of Environmental Research and Public Health	瑞士	-	253
4	BMC Public Health	英国	医学 2 区	243
5	Scientific Reports	英国	综合 3 区	175
6	Frontiers In Public Health	瑞士	医学 3 区	150
7	PLoS Neglected Tropical Diseases	美国	医学 2 区	139
8	BMC Infectious Diseases	英国	医学 3 区	138
9	Physica A Statistical Mechanics and Its Applications	荷兰	信息 2 区	113
10	Jmir Public Health and Surveillance	加拿大	医学 3 区	109

注:<sup>a</sup>:参照《中国科学院文献情报中心期刊分区表》2022 年升级版数据;<sup>b</sup>:根据“PubMed”检索截止到 2023 年 3 月 3 日数据,含提前在线发表文献;-:数据缺失

表 2 信息流行病学相关文献发表量排名前 10 位的研究机构分布(n=14 521)

序号	研究机构	国家	纳入文献量 <sup>a</sup>	纳入文献量/本研究文献总量(%)
1	University of California System	美国	458	3.15
2	University of London	英国	415	2.86
3	Udice French Research Universities	法国	369	2.54
4	Harvard University	美国	319	2.20
5	Centers for Disease Control Prevention USA	美国	292	2.01
6	World Health Organization	-	268	1.85
7	Johns Hopkins University	美国	238	1.64
8	London School of Hygiene Tropical Medicine	英国	216	1.49
9	中国疾病预防控制中心	中国	188	1.30
10	中国科学院	中国	180	1.24

注:<sup>a</sup>:非特指以第一研究机构发文;-:信息缺失

文 2 943 篇(20.27%, 2 943/14 521)(表 2)。

四、纳入研究文献的作者分析

14 521 篇文献中,统计得到全球共 63 029 位研究人员参与文献发表。其中,发文量前 10 位作者来自中国(9 位)和英国(1 位),发文量分别为 522 篇(中国,3.59%,522/14 521)、44 篇(英国,0.30%,44/14 521);前 10 位作者所在部门或单位中公共卫生(3 位)及信息技术(7 位)等相关部门的发文量分别为 143 篇(0.98%,143/14 521)和 423 篇(2.91%,423/14 521);前 10 位作者共发文 566 篇(3.90%,566/14 521)(见表 3)。

五、关键词共现及主题聚类分析

聚类可视化结果见信息流行病学研究关键词共现网络视图(图 2),共获得 3 个不同颜色的聚类主

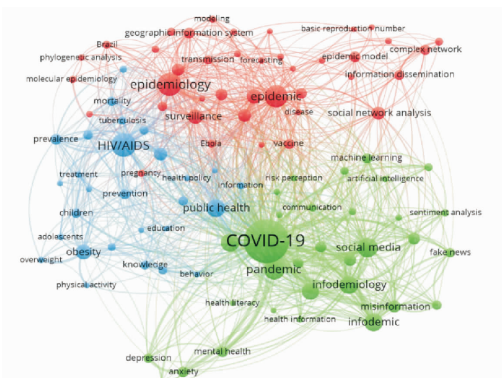
题;运用软件的可视化得到信息流行病学研究关键词标签视图(图 3),高频关键词出现的平均年份分布在 2014—2020 年,大致分为 3 个不同的时期,分别对应图 2 中的三个聚类主题。COVID-19 是至今为止信息流行病学最重要的研究领域,频次远高于第二的艾滋病(HIV/AIDS)。

聚类一(图 2 中蓝色区域):艾滋病公共卫生监测,主要涉及 28 个关键词,其中 100 频次以上关键词 9 个。与艾滋病(HIV/AIDS)关联性最强的关键词依次是监控(surveillance)、流行(prevalence)、预防(prevention)、公共卫生(public health)、分子流行病学(molecular epidemiology)等,并从图 3 可以发现该类主要关键词所对应的文献发表时间主要集中在 2014—2016 年。

表 3 信息流行病学相关文献发表量排名前 10 位的作者分布(n=14 521)

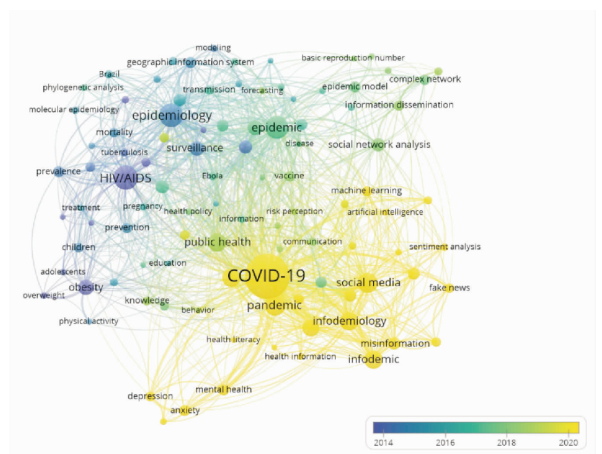
序号	作者	国家(地区) <sup>a</sup>	所在部门/单位	纳入文献量 <sup>b</sup>	纳入文献量/本研究文献总量(%)
1	Zhang Y	中国	-/电子科技大学	75	0.52
2	Wang Y	中国	数学学院/中北大学	71	0.49
3	Wang J	中国	系统工程研究院/华中科技大学	65	0.45
4	Liu Y	中国	-/重庆师范大学	61	0.42
5	Wang L	中国	电子工程学院/复旦大学	55	0.38
6	Li X	中国	-/西安交通大学附属第一人民医院	53	0.36
7	Li J	中国	信息系统与管理学院/中国人民解放军国防科技大学	48	0.33
8	Li Y	中国	电子工程学院/清华大学	48	0.33
9	Chen Y	中国	公共卫生学院/中山大学	46	0.32
10	Li L	英国	数字化公共卫生急救中心/伦敦大学学院	44	0.30

注:<sup>a</sup>:依照作者单位著录情况;<sup>b</sup>:非特指以第一作者发文;-:信息缺失



注:每种颜色代表 1 个聚类,蓝色区域代表聚类一,红色区域代表聚类二,绿色区域代表聚类三;圆球为节点标签,其越大表示关键词出现频次越多

图 2 信息流行病学研究关键词共现网络视图



注:节点颜色代表文献发表年份

图 3 信息流行病学研究关键词标签视图

聚类二(图 2 中红色区域):流行病信息监测,主要涉及 32 个关键词,其中 100 频次以上关键词 16 个。与重要领域流行病学(epidemiology)关联性最强的关键词依次是公共卫生(public health)、艾滋病(HIV/AIDS)、监控(surveillance)、建模(modeling)、流行(prevalence)、感染(infection)等;与流行病(epidemic)关联性最强的关键词依次是流行病(pandemic)、流行感冒(influenza)、公共卫生(public health)等;从图 3 发现该类关键词对应文献发表时间主要集中在 2016—2018 年左右。

聚类三(图 2 中绿色区域):COVID-19 信息流行病,主要涉及 40 个关键词,其中 100 频次以上关键词 15 个。与重要领域 COVID-19 关联性最强的关键词依次是 SARS、流行病(pandemic)、信息流行病(infodemic)、冠状病毒(coronavirus)、社交媒体(social media)、流行病(epidemic)、焦虑(anxiety)、公共卫生(public health)、精神健康(mental health)、错误信息(misinformation)、信息疫情学(infoepidemiology)、流行病学(epidemiology)等。图 3 显示发文时间主要集中于 2020 年左右。

## 讨 论

从文献年度发表数量上看,自 2000 年开始国内外信息流行病学的整体文献发表数量呈稳步上升趋势。这与 2000 年左右计算机信息等技术的普及密切相关,并且 2002 年 SARS、2009 年 H1N1 流感病毒及 2016 年 H7N9 流感病毒等都直接引起了全球对公共卫生流行病领域的关注,COVID-19 疫情使得信息流行病学研究快速发展,从 2020 开始的连续三年,每年的发表量近 2 000 篇。

在纳入文献出版期刊中,前 10 位的出版期刊所属国家来自美国、加拿大、瑞士、英国和荷兰等不同国家,累积出版量占比 14.46%。以上结果说明不同国家学术期刊对信息流行病学研究的认可度和关注度较高。前 10 位期刊以医学期刊为主(7 种),包括信息类期刊(1 种)和综合性期刊(2 种),从另一角度反映了信息流行病学是一门以医学为主的交叉性学科。

在文献发表的研究机构方面,前 10 位的机构中累计发表量居前的国家和组织分别为美国、英国、法国和中国及 WHO。一方面反映了以美国、英国等国家为代表的全球科研人员对信息流行病学研究的高度关注,全球有超过 1 万家研究机构聚焦该领域;另一方面也体现了各国研究机构在信息流行病学研究领域水平的差异。

文献发表作者信息中,排名前 9 位作者均来自中国,在一定程度上说明国内学者相对于其他国家科研人员在公共卫生相关领域,特别是流行病研究上更注重信息技术方法的应用,且国内在信息流行病学的作者主要从事电子信息工程等相关领域。综合纳入文献在出版期刊、研究机构和发表作者在医学及信息学科上的特征,反映出信息流行病学研究是一门交叉性学科。

从目前纳入文献的信息流行病学研究的关键词共现聚类结果及主题热点分析发现,2014 年至今共分为 3 个时间段,集中在 3 大热点主题,反映了公众在健康信息传播过程中其角色从被动到主动的转变。

第一阶段:2014—2016 年,该阶段信息流行病学的重要研究领域为艾滋病(HIV/AIDS),研究热点主要集中于艾滋病公共卫生监测。2014 年联合国艾滋病规划署启动了 90-90-90 目标。紧跟 WHO 研究热点,更多国内外学者从艾滋病诱因、流行病学特征、预防措施、治疗药物等内容展开监测研究<sup>[11]</sup>;不同专业学者从社交媒体、观念意识等角度研究艾滋病在公共卫生领域的防治方案<sup>[12]</sup>,有意识地针对青少年、儿童开展艾滋病防治的必要宣讲,达到艾滋病的有效防治与动态监测<sup>[13]</sup>。

第二阶段:2016—2018 年,重要研究领域为流行病学和流行病,该时期研究以流行病信息监测为主。2013 年中国首次发生了人感染 H7N9 禽流感病毒事件,2016—2017 年的大规模流行促使研究人员结合传统的流行病学方法<sup>[14]</sup>,运用信息学技术手段多角度开展相关研究。在公共卫生领域,流行病特征分析方法、传染(感染)机制、传染模型等研究内容成为主流<sup>[15]</sup>;艾滋病、H7N9 流感病毒及流感等成为

信息流行病学主要研究对象。

第三阶段:2018 年至今,围绕 COVID-19 开展了以信息流行病管理为重点的流行病学相关研究,包括 SARS-CoV-2、信息流行病、精神卫生、社交媒体等内容。与本文的研究主题信息流行病学(infodemiology) 关联性最强的关键词除前面提到四个重点研究领域的关键词外,还包括社交媒体(social media)、谷歌趋势(google trends)、信息监测(infoveillance),反映了研究领域的数据来源主要为社交媒体及谷歌趋势<sup>[10]</sup>,主要应用为信息监测。

结合信息流行病学研究热点和发展趋势,未来的信息流行病学研究还应着力解决以下几方面的问题:第一,提高研究数据的稳定性,尽力降低突发事件等因素对数据的影响,使基于数据的研究结论最大化接近实际情况;第二,拓展数据来源,扩大研究对象,目前的信息流行病学研究数据主要基于社交媒体及网络搜索趋势,覆盖的人群样本有其局限性,今后可以增加对应用软件或智能穿戴设备等数据的采集及分析;第三,进一步加强文本语义研究,目前基于深度学习或机器学习的文本语义研究,为信息流行病学的深入发展提供了契机;第四,重视数据隐私保护与伦理学研究。

由于本研究仅将 Web of Science 数据库收录文献作为代表性文献录入分析,文献稍欠全面,今后将进一步扩大权威文献数据库录入分析,为国内外相关学者全面了解并深入信息流行病学研究提供参考。

**利益冲突** 所有作者均声明不存在利益冲突

**作者贡献声明** 刘艳丽:构思设计、数据收集与整理、论文撰写;颜凡翔:数据收集与分析;吴燕萍:构思设计、数据整理与分析、论文修订

## 参 考 文 献

[1] Eysenbach G. Infodemiology: The epidemiology of (mis)information [J]. *Am J Med*, 2002, 113(9): 763-765. DOI: 10.1016/s0002-9343(02)01473-0.

- [2] Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet[J]. *J Med Internet Res*, 2009, 11(1): e11. DOI: 10.2196/jmir.1157.
- [3] Natuonal Center for Biotechnology Information. Infodemiology[EB/OL]. [2022-12-10].<https://www.ncbi.nlm.nih.gov/mesh/?term=infodemiology>.
- [4] Natuonal Center for Biotechnology Information. Infodemic[EB/OL]. [2022-12-10].<https://www.ncbi.nlm.nih.gov/mesh/?term=infodemic>.
- [5] 世界卫生组织.世卫组织首次信息流行病学会议[EB/OL]. [2022-12-10]. <https://www.who.int/zh/news-room/events/detail/2020/06/30/default-calendar/1st-who-infodemiology-conference>.
- [6] Eysenbach G. How to fight an infodemic: The four pillars of infodemic management[J]. *J Med Internet Res*, 2020,22(6):e21820. DOI: 10.2196/21820.
- [7] 吕瑾蓂,王万州,张云静,等.基于社交媒体的新冠肺炎(COVID-19)研究进展[J]. *中华疾病控制杂志*, 2020, 24(10): 1209-1214. DOI: 10.16462/j.cnki.zhjbkz.2020.10.019.
- [8] 周江杰,王胜锋,李立明. Python 爬虫技术在信息流行病学中的应用[J]. *中华流行病学杂志*, 2020, 41(6): 952-956. DOI: 10.3760/cma.j.cn112338-20190901-00643.
- [9] 刁菲,王子军. 信息流行病学在应对新冠肺炎疫情中的应用[J]. *中国公共卫生管理*, 2021,37(2):176-179.DOI: 10.19568/j.cnki.23-1318.2021.02.0010.
- [10] 周晓英,裴俊良. 信息流行病学与健康信息学[J]. *图书情报知识*, 2020(5): 126-128. DOI: 10.13366/j.dik.2020.05.122.
- [11] McNally GA. HIV and cancer: an overview of AIDS-defining and non-AIDS-defining cancers in patients with HIV[J]. *Clin J Oncol Nurs*, 2019, 23(3): 327-331. DOI: 10.1188/19.CJON.327-331.
- [12] 何慧婧,吕繁. 社会文化因素对男男性行为人群艾滋病传播的影响[J]. *中华预防医学杂志*, 2015, 49 (11): 1021-1024. DOI: 10.3760/cma.j.issn.0253-9624.2015.11.020.
- [13] 楚子君.郑州市校外青少年艾滋病知识干预效果评价[J].*郑州大学学报(医学版)*, 2015, 50(6): 854-856. DOI:10.13705/j.issn.1671-6825.2015.06.031.
- [14] 孙海燕,童海江,崔大伟,等. 人感染 H7N9 禽流感病毒的研究进展[J]. *中华临床感染病杂志*, 2017, 10 (1): 68-75. DOI: 10.3760/cma.j.issn.1674-2397.2017.01.015.
- [15] Kirby RS, Delmelle E, Eberth JM. Advances in spatial epidemiology and geographic information systems[J]. *Ann Epidemiol*, 2017, 27(1):1-9. DOI: 10.1016/j.annepidem.2016.12.001.
- [16] Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data[J]. *Nature*, 2009, 457 (7232):1012-1014. DOI: 10.1038/nature07634.

(收稿日期:2022-12-14)