

·医学信息学·

2015- 2020

郑津津,张惠荣,黄育北

(天津医科大学肿瘤医院图书情报研究室/国家肿瘤临床医学研究中心,天津 300060)

**摘要:**目的 分析近5年精准医学高被引论文的研究状况和发展动态。**方法** 基于Web of Science核心合集数据库,利用Citespace、文献计量在线分析平台等分析软件对2015-2020年我国精准医学研究的高被引文献进行计量分析。**结果** 共纳入350篇文献,2015-2020年我国精准医学的发文章量呈较快的增长趋势;与中国精准医学合作最多的国家/地区是美国、英国与德国;收录高被引论文具有影响力的期刊分别为《NATURE REVIEWS DRUG DISCOVERY》《LANCET》和《NATURE REVIEWS CLINICAL ONCOLOGY》;研究高频关键词分别为 Precision medicine、Cancer、Personalized medicine、Breast cancer 和 Medicine。**结论** 目前我国精准医学研究内容逐渐面向国家政策、规范质量监督与管理、临床实践诊疗指南路径等相关领域倾斜,体现了精准医学的宏观发展。

**关键词:**精准医学;高被引论文;知识图谱;诊疗指南

**中图分类号:**R73;G250.252

**文献标识码:**A

**DOI:**10.3969/j.issn.1006-1959.2021.24.003

**文章编号:**1006-1959(2021)24-0009-06

### Analysis of Highly Cited Papers in Precision Medicine from 2015 to 2020

ZHENG Jin-jin,ZHANG Hui-rong,HUANG Yu-bei

(Medical Library,Cancer Institute & Hospital of Tianjin Medical University/National Cancer Clinical

Medicine Research Center,Tianjin 300060,China)

**Abstract: Objective** To analyze the research status and development trend of highly cited papers in precision medicine in recent 5 years.  
**Methods** Based on the core collection database of Web of Science, the highly cited literature of precision medicine research in China from 2015 to 2020 was analyzed by using the analysis software such as Citespace and bibliometric online analysis platform.  
**Results** A total of 350 articles were included, and the number of published articles on precision medicine in China showed a rapid growth trend from 2015 to 2020. The countries/regions that had the most cooperation with precision medicine in China were the United States, Britain and Germany; the influential journals included in highly cited papers were *NATURE REVIEWS DRUG DISCOVERY*, *LANCET* and *NATURE REVIEWS CLINICAL ONCOLOGY*; the high frequency keywords were Precision medicine, Cancer, Personalized medicine, Breast cancer and Medicine.  
**Conclusion** At present, the research content of precision medicine in China is gradually oriented to national policies, standardized quality supervision and management, clinical practice, diagnosis and treatment guidelines and paths, which reflects the macro development of precision medicine.

**Key words:** Precision medicine; Highly cited papers; Knowledge map; Clinic guide

在国际精准医学发展计划的背景下,我国于2015年2月初首次提出“精准医学”概念,3月11日建立了精准医学的研究专家组,开展该计划,通过不断的努力,促进了精准医学计划的发展<sup>[1]</sup>。2016年初,在“十三五”发展规划中,正式启动了对精准医学的研究,制定了相应的政策和发展规划,同时将它列为我国重点发展计划之一,随之全国范围内的不同机构,迅速开展了针对不同疾病的“精准医学”研究,极大的促进了其发展<sup>[2]</sup>。精准医学(precision medicine)是以个体化医疗为基础,随着分子生物学和高通量基因测序技术的飞速发展以及生物信息与大数据科学的交叉应用而发展起来的医疗模式<sup>[3,4]</sup>。关于精准医学的文章数量迅速增长,表明精准医学的探究正在迅速发展。本文基于文献与引文分析概述了精准医学发展的演化轨迹,并强调该学科发展的历程。基于CiteSpace选择的网络结构和计算属性分析发现学科发展趋势和模式,旨在促进该领域文

献的科学前沿与发展动态<sup>[5]</sup>。视觉分析工具对于发现大量新发表研究中的关键发展节点具有较高研究价值,通过跟踪引文链接和引用矩阵间的相互关系,科学计量学的视觉分析技术能够引导学者关注到振动和迅速发展的研究前沿与热点。本研究基于Web of Science核心合集数据库,以高被引SCI论文数、影响因子、作者/机构合作关系、高频关键词为定量评价指标,运用CiteSpace、文献计量在线分析平台从文献计量学角度对2015-2020年精准医学高被引论文进行研究,旨在评价精准医学的发展态势,为我国制定精准医学学科发展规划提供参考依据。

#### 1 资料与方法

**1.1 资料来源** 基于Web of Science核心合集(Science Citation Index Expanded)和2019版Journal Citation Report(JCR),以“Precision medicine”作为检索主题词,数据采集时间为2015年1月1日-2020年12月31日,数据更新时间为2021年3月31日,得到14 237篇基础数据,依据领域中的高被引论文过滤结果,最终获得有效数据为350篇,纳入本研究。

**1.2 纳入与排除标准** 纳入标准:①检索地址中包含精准医学相关研究文献;②满足符合文献计量学所

基金项目:国家自然科学基金青年项目(编号:81502476)

作者简介:郑津津(1984.2-),女,天津人,硕士,馆员,主要从事数据挖掘与文献计量与科学评价研究

需的相关高被引文献。排除标准:①涵盖信息不全的文献;②重复或撤稿的文献;③会议论文、社论、书籍章节。由两位作者独立对符合条件的文献进行凝练、提取及开展数据分析。

1.3 方法 采用软件 BibExcel 制作发文量曲线。采用文献计量分析平台研究国家与机构合作关系及期刊影响力分析,采用 CiteSpace 5.7.3 软件进行关键词聚类分析、被引文献聚类、bursts(突发词)检测等知识图谱的绘制。

## 2 结果

2.1 文献总发文量分析 高被引论文数量与精准医学的发展有着直接且必然的联系,精准医学领域高被引论文于 2015 年发文量为 36 篇,2016 年发文量为 47 篇,2017 年发文量为 56 篇,2018 年发文量为 73 篇,2019 年发文量为 86 篇,但需要注意的是,由于数据截至 2020 年底,导致 2020 年的数据情况未

能全部体现,仅为 52 篇,见图 1。

### 2.2 合作关系分析

2.2.1 国家间合作关系分析 当前,美国通常从探究肿瘤靶向治疗的创新方法并在临床水平上展开实践,采取综合疗法、克服肿瘤的耐药性等方面进行肿瘤的诊断和治疗。弗吉尼亚大学采用基因组信息创建电子健康病历,进行风险评估,对提供者和患者给予决策支持;杜克大学开发评估肺癌和乳腺癌药物基因组学编织物的支持证据生成和临床疗效评价研究;H Lee Moffit 癌症中心建设 CER 信息基础数据集;Fred Hutchinson 癌症中心开展决策建模、伦理以及开展临床试验涉及等研究。美国国立卫生院前院长认为应该将基础研究与临床实践紧密联系,这一观点对于医学的发展具有里程碑意义<sup>[6]</sup>。各国家间合作关系见图 2。

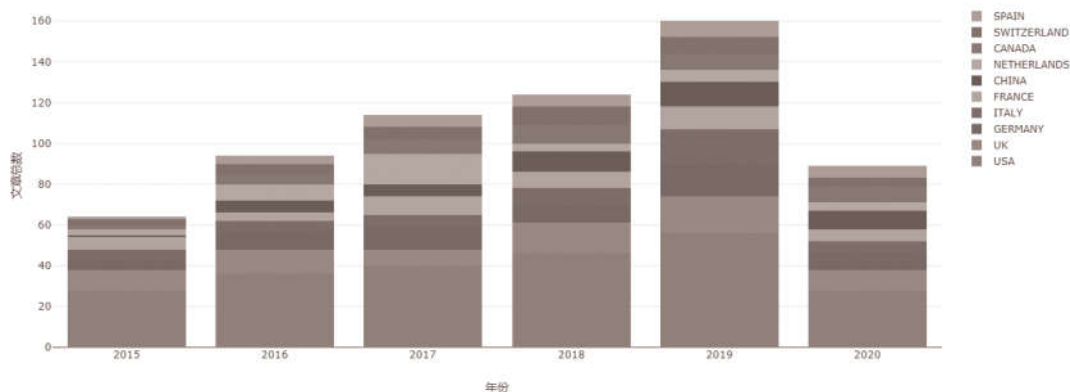


图 1 历年发文量之国家发文量

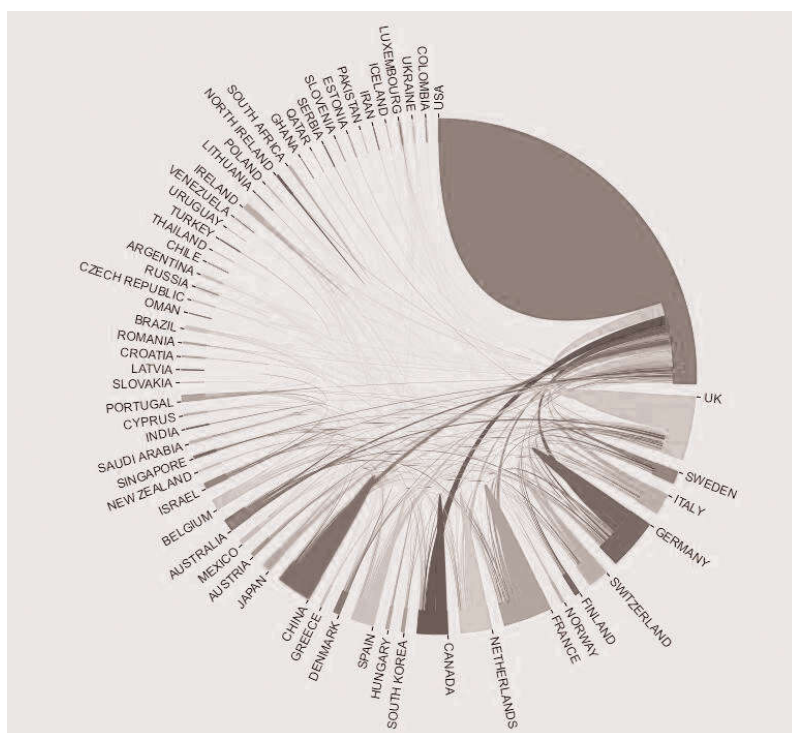


图 2 国家间合作关系

2.2.2 机构间合作关系分析 精准医学的发展和科研产出机构有着紧密的关联性。在对产出机构(机构包括大学,各研究院,以及国家研究机关)的学术位置进行评价时,将中心性以及所在节点相结合,可将研究的学术状况以及影响力较好地展现出来。各机

构间合作关系见图 3。  
2.3 期刊影响力分析 结合文献计量分析平台,汇总 2015–2020 年精准医学高被引论文最具影响力期刊,见表 1。

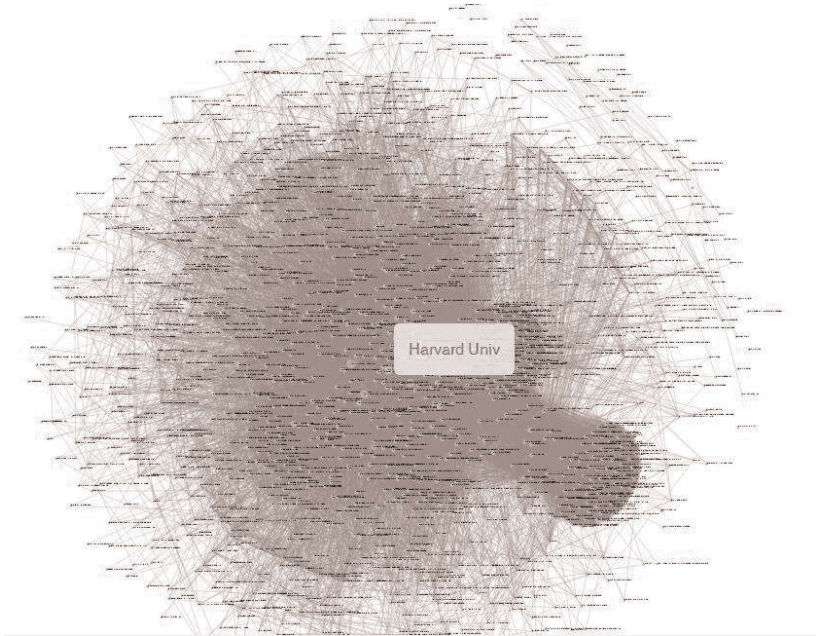


图 3 机构间合作关系

表 1 2015–2020 年精准医学高被引论文最具影响力期刊 Top10

序号	期刊名	2019 版影 响因子	JCR 类别
1	CANCER DISCOVERY	29.497	ONCOLOGY
2	NATURE REVIEWS CANCER	53.03	ONCOLOGY
3	NATURE	42.779	MULTID-ISCIPLINARY SCIENCES
4	NATURE REVIEWS CLINICAL ONCOLOGY	53.276	ONCOLOGY
5	ANNALS OF ONCOLOGY	18.274	ONCOLOGY
6	CELL	38.637	BIOCHEMISTRY & MOLECULAR BIOLOGY
7	NATURE REVIEWS DRUG DISCOVERY	67.797	BIOTECHNOLOGY & APPLIED MICROBIOLOGY
8	LANCET	60.39	MEDICINE, GENERAL & INTERNAL
9	NATURE MEDICINE	36.13	BIOCHEMISTRY & MOLECULAR BIOLOGY
10	CANCER CELL	26.602	CELL BIOLOGY

序号	收录高被引 文章总数	收录高被引 文章占比	总被引用 次数	平均被引 次数
1	11	3.143%	17	1.55
2	9	2.571%	14	1.56
3	9	2.571%	5	0.56
4	8	2.286%	2	0.25
5	7	2.000%	4	0.57
6	6	1.714%	11	1.83
7	5	1.429%	5	1.00
8	5	1.429%	0	0.00
9	5	1.429%	0	0.00
10	5	1.429%	0	0.00

2.4 关键词聚类分析 我国关于“精准医学”的研究快速增长,从 2015 年至今,5 年内精准医学相关文献的发表量超过了最初发表量的 5 倍。从 WOS 数据库年度发文数量来看,因数据库存在局限性,国际精准医学的研究先于我国在 1998–2015 年开始,但发文量很少。自 2016 年起,国际上此方面的研究持续增多,且趋向于快速增长,在发展速度上远胜于我国。对 WOS 数据库下载的数据进行聚类分析,主要包括作者、题目、摘要和文献的引文等字段,即研究所要抽取的文本数据项。输入至系统后,设置参数,如确定要分析领域的总时间段范围、分割后每一个时间片段的长度,本文的时间段范围是 2015–2020 年,时间片段的长度是每 1 年一个,术语来源选择 Title (标题)、Abstract (摘要)、Author Keywords DE (作者关键词)、Keywords Plus Identifiers (关键词标识项),阈值选择 Top50,节点类型确定为 Keyword,

然后根据不同的研究目的(研究领域探析、研究前沿探析)的需要选取不同的选择项:名词短语(Noun phrase)和突变词(Burstterm)首次运行取消术语类型的选择,再根据需要设定一定的阈值以及路径选择算法(Pathfinder),其他选项保持默认状态。与学科研究进展相关的关键词见表 2。国际精准医学前沿的高 Burst 值关键字排名前 10 位见图 4。

文献的关键词共现分析通常多用以对某个研究领域的热点与前沿问题进行清晰明确研究,一般会将突现率较高的关键词视为精准医学前沿领域的表征<sup>[7]</sup>。为进一步研究精准医学领域的研究热点,采用 CiteSpace 构建了主题词加关键词构成的共现网络,用以微观层面揭示学科内的实体关系特征,构建知识单元之间直接的语义关联,并从施引文献的标题中提取聚类命名术语,经整理生成关键词被引频次 $\geq 20$ 的关键词列表,关键词聚类分析见图 5。

表 2 学科研究进展关键词

序号	关键词	年份	中介中心度	文章数量
1	Precision medicine	2015	0.15	339
2	Cancer	2018	0.15	307
3	Personalized medicine	2015	0.13	216
4	Breast cancer	2016	0.12	198
5	medicine	2018	0.15	179
6	Biomarker	2017	0.10	168
7	Drug therapy	2019	0.12	155
8	Clinical trail	2020	0.16	135
9	Cell lung cancer	2020	0.12	121
10	Gene	2019	0.07	109



图 4 排名前 10 位的国际精准医学前沿的高 Burst 值关键字



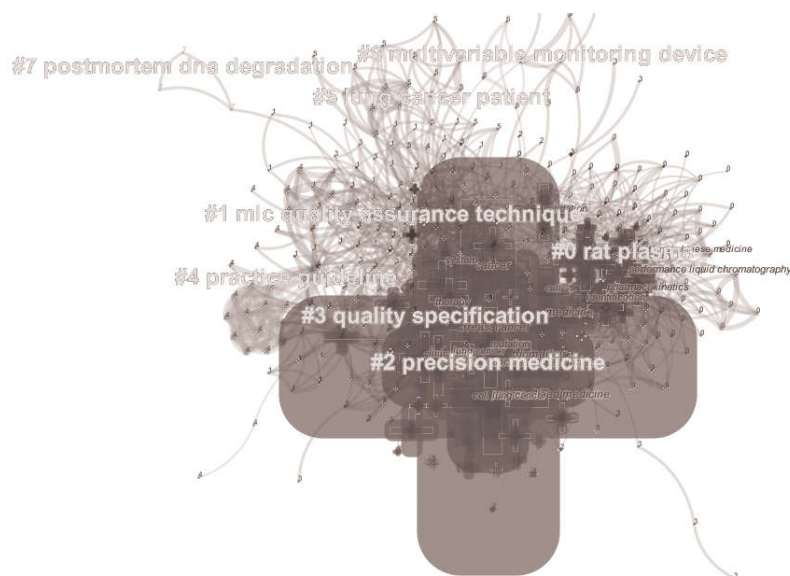


图 5 关键词聚类分析

### 3 讨论

中国有无可比拟的病患样本资源优势,本文通过对近 5 年精准医学领域高被引论文的期刊影响力分析、作者/机构间生产力与合作关系、领域内的关键词及研究热点聚类分析,开展文献的梳理与归纳总结,结合定性的文献回顾和定量的文献计量学分析方法,得出主要研究结论如下。

**3.1 引文总量情况** 作为“文献计量”数据来源的学术论文是科学研究的主要对象之一,论文被引频次的高低,在一定程度上可以揭示论文的影响程度和情报价值<sup>[8,9]</sup>。高被引论文是指被引用频次相对较高,被引用周期相对稳定的研究论文<sup>[10]</sup>。从 2015 年至 2019 年,各国精准医学高被引论文整体呈上升趋势,尤其在 2019 年达到高峰,占总发文量的 25%。我国继美国、英国、德国、意大利、法国等欧美国家后排名第六位,占比 14%,呈较快的发展态势。

**3.2 研究机构情况** 科研知识产出能力、学术影响力可分别借助其合作网络分析进行展现<sup>[11]</sup>。在国家间合作关系中,2015–2020 年,中国和 58 个国家与地区进行国际合作发表精准医学高被引 SCI 论文,其中美国是合作发文排名第一的国家与地区共计 236 篇,占总文章量 67.429%;英国是合作发文排名第二的国家与地区共计 70 篇,占总文章量 20%,德国是合作发文排名第三的国家与地区共计 59 篇,占总文章量 16.857%。据统计,美国则拥有最高的相对合作强度,其次是英国、德国。中国精准医学研究领域处于快速发展阶段,我国学者应加大与国际他国之间的高水平合作,进一步推动我国精准医学国际化的战略合作关系。

借助对精准医学高被引论文学术影响力较大的产出机构的追踪,得出 HARVARD UNIVERSITY 所

发表的文章位列首位,共计 71 篇,占比 20.286%,其次是 UNIVERSITY OF CALIFORNIA SYSTEM,共计 46 篇,占比 13.143%,再次是 UNIVERSITY OF TEXAS SYSTEM,共计 35 篇,占比 10%,文章中涉及到精准医学基于多组学获取的临床试验数据及生物信息学技术用于临床诊断及精准医学所面临的主要挑战和基于数据集成方法的最新进展,以揭示各种组学研究产生的大数据中的个性化信息等,上述机构汇聚了此研究领域内的核心及重点内容,其得出之结论、所用之方法或许会对后期的研究产生一定的影响。同时,从研究视角发现应加大数据资源管理与控制能力,包括对大数据资源的质量监管,提升数据处理的准确性、实时性、有效性,加强数据资源的利用和隐私保护,积极开展基础队列研究、疾病基因检测技术、精准医学知识库建设及多方大数据平台共享机制的建设<sup>[12]</sup>,提升挖掘技术的开发能力,整合全生命周期电子档案系统,建立符合中国模式的精准医学大数据研究平台和保障体系,培养具有精准医学发展需求的高能人才梯队建设,培养医学信息学专家、科学项目管理人才及医学伦理学人才作为支撑,制定属于中国精准医学的疾病诊断治疗标准、疾病预防标准,进一步提升医疗的社会普惠性,使重大病症的发病率、死亡率、临床治疗费用大大降低,促进我国医疗事业的快速发展。

**3.3 最具影响力期刊情况** 学术期刊的学术辐射力和市场占有率是学术期刊影响力的本质特征<sup>[13]</sup>。近 5 年间精准医学高被引论文最具影响力 Top10 期刊,依据 2019 版 JCR 影响因子排在前 3 位的期刊依次为《NATURE REVIEWS DRUG DISCOVERY》,IF=67.797;《LANCET》,IF=60.39;《NATURE REVIEWS CLINICAL ONCOLOGY》,IF=53.276。总影响

因子之和为426.392,平均影响因子为42.6392,涉及主要学科领域分别为肿瘤学(ONCOLOGY)和生物化学与分子生物学(BIOCHEMISTRY & MOLECULAR BIOLOGY),这也可以从一个侧面反映出精准医学在该学科的发展态势以及研究的重点方向。收录高被引文章总数高达11篇,总被引用次数高达17次,为精准医学在肿瘤学领域的传播和发展起到了重要的引领作用。

3.4 研究热点情况 关键词是作者对论文核心研究内容的精炼<sup>[13]</sup>,聚类分析则是一种多元统计法,主要是完成对涵盖有诸多指标数据的有效分类,即基于层级聚类的原则,对于事物彼此间的亲密及疏离程度的判定借助相似性尺度来完成,以清晰展现出文献彼此间的亲疏关联性<sup>[14,15]</sup>。基于CiteSpace软件分析得出,Precision medicine 频次最高,其次是Cancer、personalized medicine、Breast cancer、Medicine、Biomarker、Drug therapy、Clinical trial、Cell lungcancer和Gene等高频关键词。以CiteSpace对Web Of Science核心合集的数据展开Keyword值的Burst分析,所得出的Burst值超过3.53的关键词共有21个,其中中介中心度>0.1的数值最具参考价值。

我国精准医学的发展呈快速发展阶段,还需结合近几年的科研成果及临床经验,以基因组研究为代表的各种分子水平的资源与信息技术,通过挖掘与整理,整合出能够和疾病相关的且能够作为疾病表征的标本组织库<sup>[17,18]</sup>。科研决策机构应加强政策引导,在制定年度计划和战略目标时,从学科建设入手,联合生产基因检测产品的医疗器械公司及药物公司,在市场前景和优惠政策的激励下,增加相关企业数量,借助先进的计算机技术和AI技术,提升非结构化数据的处理效率,增强数据的准确性、可靠性,促进精准医学行业的飞速进步。充分发挥所在单位研究院所、临床医学研究中心在基础与临床研究相结合的核心作用,促进科研工作者的积极研究,形成一套具有我国自主专利、获得全球认可的治疗指南和临床经验路径,极大的提升治疗重大疾病的医疗水平。

基于以上研究,本文认为中国与国际前沿技术和理念接轨的同时,仍然需建立国际一流的精准医学研究平台、保障体系、自主研发的核心关键技术、适合亚洲种群的防疫药物、疫苗、器械和设备等,建立与制定人类指南、临床路径和干预措施,应用于疾病的预防、分析、诊断、治疗、预后的各个阶段,尤其在治疗肿瘤、糖尿病等恶性疾病中,研制多种靶向药物设计方案<sup>[19]</sup>,在全国范围内进行推广,极大提升治疗效果,为广大患者带来福音,同时还将更加具有可塑性和可信性,从而爆发出巨大的医疗经济潜能,推动我国总体医疗的快速发展。

## 参考文献:

- [1]Stephanie Devaney.Nominate a White House Champion of Change for PrecisionMedicine[EB/OL].[2016-03-17]. <https://www.whitehouse.gov/blog/2015/05/18/nominate-white-house-champion-change-precision-medicine>.
- [2]国家自然科学基金委员会管理科学部.管理科学发展战略:暨管理科学“十三五”优先资助领域[M].北京:科学出版社,2016.
- [3]吴斌,占美,徐琰,等.我国精准医学概念的循证研究[J].中国药房,2017,28(8):1017-1022.
- [4]Lyman GH,Moses HL.Biomarker Tests for Molecularly Targeted Therapies: Laying the Foundation and Fulfilling the Dream[J].J Clin Oncol,2016,34(17):2061-2066.
- [5]刘亚静,潘云涛,赵筱媛,等.精准医学领域的研究现状及其可视化分析[J].中华医学图书情报杂志,2017,26(8):41-49,53.
- [6]刘哲然.从经验医学、循证医学到精准医学的演变及评价[J].医学与哲学,2017,38(10):81-84.
- [7]康宏宇,李姣,刘燕,等.精准医学研究相关信息资源的发现与汇聚[J].中华医学图书情报杂志,2017,26(11):1-6.
- [8]赵鸿玉,薛丽娜,王代礼.国内外LIS领域高被引论文比较研究[J].情报科学,2021,39(4):129-139.
- [9]黄晓斌,张欢庆.我国情报学高被引论文分析[J].情报科学,2018,36(1):54-60.
- [10]Olisah C,Okoh OO,Okoh AI.A bibliometric analysis of investigations of polybrominated diphenyl ethers (PBDEs) in biological and environmental matrices from 1992-2018[J].Heliyon,2018,4(11):e00964.
- [11]刘作仪,吴登生,李建平.2001-2010年我国管理与运筹学研究态势的计量分析——基于Web of Science数据[J].北京理工大学学报(社会科学版),2012(1):1-8.
- [12]Jacob MP,Thomas J,Durant S,et al.Health Care and Precision Medicine Research: Analysis of a Scalable Data Science Platform[J].J Med Internet Res,2019,21(4):e13043.
- [13]何汶.中国学术期刊影响力评价分析[J].重庆大学学报(社会科学版),2008(2):85-89.
- [14]朱明,谢梦晴,刘宇.近十年国内图书馆学研究热点述评:基于高被引论文的计量分析[J].高校图书馆工作,2019,39(1):5-10,54.
- [15]范美玉,陈敏.基于大数据的精准医疗服务体系研究[J].中国医院管理,2016,36(1):10-11.
- [16]徐速,李维.精准医学研究热点的双向聚类计量分析[J].医学与哲学,2015,36(6):1-5,34.
- [17]孙小康.精准医学大数据汇交与共享政策研究[D].北京:北京协和医学院,2018.
- [18]王宇,王心慰,刘爽,等.精准医学的临床部署:顶层架构设计及关键信息技术[J].转化医学杂志,2017,6(6):321-324.
- [19]du Rusquec P,Filleron T,Le Tourneau C.Drug development in the era of precision medicine in oncology-Clues beyond randomization[J].Chin Clin Oncol,2021,10(3):23.

收稿日期:2021-05-20;修回日期:2021-06-23

编辑/成森