

基于百度指数的乳腺癌防治健康信息网络关注特征分析*

朱笑笑 钱爱兵

(南京中医药大学卫生经济管理学院 南京 210023)

[摘要] 利用百度指数获取乳腺癌诊治健康信息的网民搜索及媒体报道数据,从时间分布、群体特征、相关搜索词、头条新闻内容及来源等维度分析公众网络关注特征。结果显示网民关注乳腺癌早期症状及预后效果,关注度随乳腺癌发病率提高而增加。媒体多发布乳腺癌致病因素、治疗方式、预防保健相关新闻,报道主题与乳腺癌防治的突破性进展相关。

[关键词] 乳腺癌;健康信息;百度指数;信息需求

[中图分类号] R-056 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2020.03.006

Analysis of Network Attention Characteristics of Breast Cancer Prevention and Treatment Health Information Based on Baidu Index

ZHU Xiaoxiao, QIAN Aibing, School of Economics and Health Management, Nanjing University of Chinese Medicine, Nanjing 210023, China

[Abstract] By using Baidu index to obtain data on the netizens' search and media coverage of breast cancer diagnosis and treatment health information, the paper analyzes the characteristics of public network attention from dimensions of time distribution, group characteristics, related search terms, contents and sources of top news, etc. Results show that netizens pay attention to the early symptoms and prognosis effect of breast cancer, and the attention increases with the incidence of breast cancer. The media often release news about breast cancer pathogenic factors, treatment methods, prevention and health care, and the topics covered are related to the breakthroughs in breast cancer prevention and treatment.

[Keywords] breast cancer; health information; Baidu index; information requirement

[收稿日期] 2019-07-16

[作者简介] 朱笑笑,硕士研究生,发表论文24篇;通讯作者:钱爱兵,博士,副教授,硕士生导师。

[基金项目] 江苏省社会科学基金项目“健康中国视野下健康信息的组织与利用模式研究”(项目编号:18TQA003);江苏省教育厅高校哲学社会科学基金项目“江苏省医药卫生期刊影响力研究”(项目编号:2016SJB870007);江苏省研究生科研与实践创新计划项目“基于新媒体的乳腺癌健康信息传播及用户搜寻行为研究”(项目编号:KYCX19_1243)。

1 引言

互联网使公众能随时随地获取健康信息。截至2018年底我国搜索引擎用户达6.81亿^[1],63.16%的搜索主题为健康与医疗^[2]。分析信息行为有助于把握网民特征及社会关注点。龚韩湘等^[3]基于百度指数认为公众最关注居家养老;王云娣^[4]发现网民及媒体对公共文化服务体系的关注度受政府驱动影响;李清光^[5]得出食品安全事件关注度与网民信息行为相关的结论。本研究利用百度指数获取乳腺癌防治健康信息的网民搜索及媒体报道数据,从时间

分布、群体特征、相关搜索词、头条新闻内容及来源等维度分析公众网络关注特征。

2 资料与方法

2.1 资料来源

百度指数通过计算关键词在搜索引擎中出现频次的加权和,反映公众对特定主题的关注度。以“乳腺癌”为检索词,获取2011-2018年中国大陆地区31个省份(不含港澳台)的网民在PC及移动端对乳腺癌防治健康信息的日均搜索频次及媒体日均报道频次,以《中国肿瘤》2011-2015年的乳腺癌粗发病率数据^[6-10]反映疾病发展规律。

2.2 分析过程

使用数据挖掘方法并借助Excel处理数据。以搜索指数反映网民信息行为,通过CORREL函数分析搜索指数与我国乳腺癌粗发病率的相关性。基于人群画像分析网民群体特征,地域分布的观察期为2013年7月-2018年12月,性别、年龄分布的观察期为2013年9月-2018年12月。采集2018年每周排名前10位的相关检索词,共获得520个来源词及520个去向词,过滤、合并、分类统计高频检索词。以媒体指数反映媒体关注度及信息曝光度,描述分析乳腺癌相关新闻的年度变化趋势。基于峰值统计2013年9月-2018年12月的头条新闻,去除无关数据,共获得21个季度的578条新闻,人工标注15个主题。去除失效链接,522条新闻来自139个网站,对信息源进行分类汇总。

3 乳腺癌防治健康信息网民关注特征

3.1 时间分布

百度搜索指数映射网民信息搜索频率,粗发病率反映一定时间内特定人群的癌症实际发生情况。其趋势,见图1。乳腺癌防治健康信息的网民搜索频次与其粗发病率在特定时段变化一致,二者相关系数为0.89。乳腺癌发病率升高,网民信息搜索行

为频繁,对乳腺健康的重视度提高。2015年1月16日歌手姚贝娜因乳腺癌复发去世,引起网民对乳腺健康的讨论。15日搜索值仅为12 242,16日暴涨至313 389,且热度持续至23日。名人效应使公众在短时间内聚焦乳腺癌并基于网络获取健康信息。

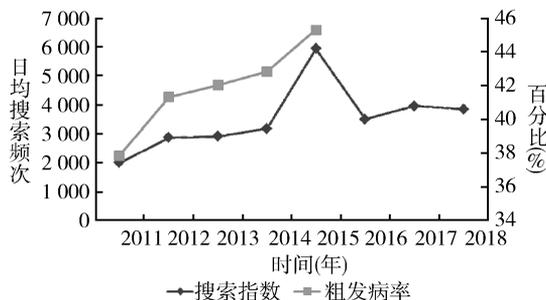


图1 乳腺癌粗发病率及网民搜索指数变化

3.2 网民群体

3.2.1 地域 广东省网民搜索乳腺癌防治健康信息的频次最高,江苏、浙江其次。排名前10位省市搜索指数年度变化,见图2。去除峰值影响,四川省搜索指数在2017年下降10.74%,次年下降4.01%,河北省在2018年下降3.99%,其余省市搜索指数均逐年升高但涨幅减小。观察期内高搜索指数省份乳腺癌的发病率、检出率或死亡率均较高或处于持续上升状态,乳腺癌为该地区女性群体的主要恶性肿瘤。80%的高搜索指数省份位于东部地区。该区域经济发展快且城市化水平高,可能导致生育率下降、母乳喂养减少、高脂肪饮食增加等致癌因素增加,东部地区网民更关注乳腺健康。

3.2.2 性别 乳腺癌虽为女性高发肿瘤,也存在男性病例,乳腺健康应为全民关注话题。网民性别比变化,见图3。2013-2015年60%的搜索用户为男性,2016-2018年女性为信息搜索主要群体,男女比约为3:7。而我国全体网民男女比为53.7:47.3^[1],乳腺健康信息搜索行为存在性别差异。观察期内女性用户持续增加,乳腺癌发病率升高使大量女性对乳腺健康相关建议的需求增加,通过网络匿名化获取健康指导。前期男性占比较高说明该群体也愿意主动了解乳腺健康知识。

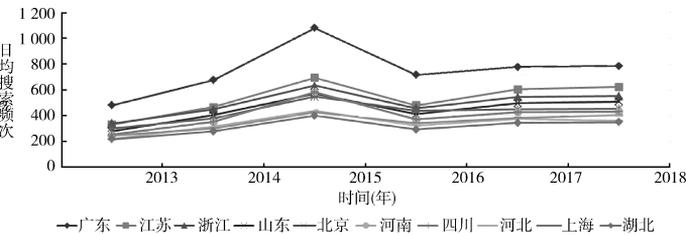


图2 各省市搜索指数年度变化

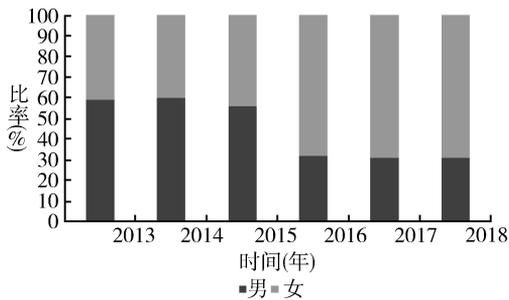


图3 网民群体性别比变化

3.2.3 年龄 乳腺癌发病率与年龄相关, 统计各年度年龄比变化以分析不同群体的信息搜索行为偏好, 见图4。搜索乳腺健康知识的网民存在年龄分层, 29岁以下用户占比降低, 30~39岁用户为主要群体, 40~49岁占比上升, 50岁以上保持在6%左右。我国网民以20~29岁群体为主, 30~49岁占39.1%, 50岁以上已上升至12.5%^[1]。对比可知30~49岁网民更愿意搜索乳腺健康信息。30岁后乳腺癌发病率随年龄增长快速增加, 57.4%的患者在50岁前发病^[11]。高发病率的中青年女性对乳腺保健更关注, 信息搜索行为也较频繁。50岁以上用户较少可能是因为搜索引擎不是该群体获取健康信息的主要渠道。

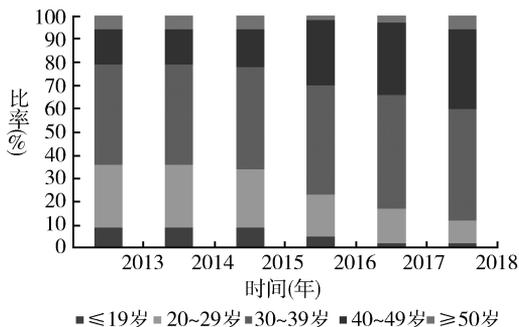


图4 网民群体年龄占比变化

3.3 相关搜索词特征

来源相关词反映用户在搜索乳腺癌前的信息偏好, 去向相关词体现在获取相关健康知识后用户的信息需求。高频相关检索词, 见表1。“乳腺癌的早期症状”在来源及去向词中均排第1位。与早期症状相关的来源词占36.54%, 乳腺癌的早发现、早诊断、早治疗是现阶段最有效的防治手段。“增生”、“乳腺结节”、“囊肿”等症状刺激网民继续搜索防癌健康指导。“宫颈癌”、“宫颈”频次较高, 宫颈癌及乳腺癌为我国“两癌筛查”对象, 在进行宫颈癌防治健康信息搜索时, 用户可能收到乳腺健康信息的推送, 促使其检索乳腺癌主题。从“怎么回事”、“多少岁”等疑问词可以推断现阶段乳腺健康知识普及度及全面性较低。高频去向相关词总结网民在了解乳腺癌基本信息后继续搜索的主题范围。“乳腺癌的早期症状”、“早期”等词高频出现, 体现网民偏好借助网络信息自检自查。“能治好吗”、“晚期”、“术后”、“能活几年”反映网民对乳腺癌的恐惧心理, 迫切想要了解其治疗手段、生存年限等。“真人图片”反映大部分网民对相关术语、专业表述的理解程度较低, 需要通过图片辅助理解。对比发现来源词中乳腺癌的早期症状等信息占比最高, 去向词中治疗手段、预后效果等主题频次上升。检索词变化反映网民感知健康信息的过程: 根据某些症状怀疑患有某类疾病——查阅疾病早期特征判断患病可能性——初步了解疾病治疗手段及预后效果。

表1 乳腺癌高频相关检索词

序号	来源相关词	频次	去向相关词	频次
1	乳腺癌的早期症状	50	乳腺癌的早期症状	86
2	增生	49	早期	52
3	早期	42	症状	51
4	症状	40	乳腺癌是怎么引起的	49
5	宫颈癌	37	增生	36
6	宫颈	33	乳腺癌能治好吗	31
7	乳腺结节	44	晚期	22
8	乳腺癌是怎么引起的	13	术后	19
9	乳腺囊肿是怎么回事	5	乳腺癌症状真人图片	18
10	乳腺癌最早多少岁	2	乳腺癌手术后能活几年	16

4 乳腺癌防治健康信息媒体关注特征

4.1 时间分布

百度媒体指数反映网络媒体报道中与乳腺癌相关且被百度新闻收录的数量,其趋势,见图5。媒体对乳腺癌的关注度存在年度差异,且整体波动下降。2015年媒体关注度达到峰值,姚贝娜因乳腺癌复发去世的消息引起社会轰动,媒体高度关注并报道较多相关新闻,当日媒体指数环比增长17倍,次日热度逐渐降低。近几年媒体指数保持较低稳定值,乳腺癌相关新闻的报道频次较低。可能原因是媒体报道力求信息新颖性及轰动效应,而乳腺癌为女性高发恶性肿瘤已是常识,且近几年其疾病特征未发生变化,防治、预后效果也未取得突破,媒体仅保持较低曝光率。

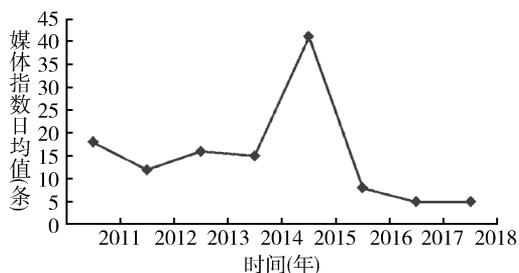


图5 乳腺癌健康信息媒体关注度年度变化

4.2 头条新闻

4.2.1 内容 头条新闻即观察期内最具社会影响力的报道,高频新闻主题及主要内容,见表2。乳腺癌致病因素及预防保健为媒体关注重点。吸烟、饮酒、肥胖、精神压力均会增加乳腺癌发病率,媒体通过新闻提醒公众改善生活习惯并保持身心健康。46.10%的预防保健知识为饮食防癌,包括食用海带、豆制品、西红柿、石榴等。10.11%的预防措施为运动,包括乳房保健操、适度锻炼150分钟等。仅有两条新闻指出早筛早诊早治在乳腺癌预防中的重要作用。研究进展及治疗方式的曝光度也较高,媒体主要关注新药研发、新的转移通路及病变基因,以及治疗方式如术中放疗、内分泌治疗、膏方调养及微创手术等。特

殊患者的高频报道体现媒体对男性、名人及低龄患者较为敏感,希望通过特殊案例引起公众对乳腺健康的重视。疾病特征主要介绍乳腺癌发病率及死亡率变化,如“发病率排女性恶性肿瘤首位”。诊断检测涉及血检、尿检等多种早期检测方法,提示公众通过体检实现乳腺癌的早发现及早治疗。病例报道主要宣传积极乐观的患者形象以提高患者生存意愿、社会对乳腺癌的关注度及助力等。公益活动多是借助明星宣传或社会资助等方式宣教乳腺癌防治知识。专家辟谣类新闻的出现是由于网络信息冗杂造成不实信息大范围传播,如豆浆、蜂王浆致癌,造成公众恐慌。行业发展、自检自查等为小众主题,包括药企融资、健康指南、产品广告、重疾赔付等。新闻多样性表明媒体关注乳腺癌防治、患者群体等全方位信息,通过报道疾病诱因、食疗预防、治疗进展等新闻,采用夸张、自问自答等形式吸引公众注意。进一步统计每年各主题相关头条新闻,分析媒体报道主题偏好,主题词年度变化,见图6。2013年媒体对疾病特征的曝光率最高,报道集中在疾病特征类信息,“乳腺癌成为女性第一大癌”、“我国乳腺癌发病率提早”等新闻占据头条。2014年疾病特征类头条新闻仅9条,居第5位,致病因素类信息跃居首位。“吸100支烟患乳腺癌危险增加三成”、“久坐不动26岁女白领患上乳腺癌”等新闻披露乳腺癌的可能致病因素。2015年致病因素主题新闻17条,预防保健类16条,媒体通过“乳腺癌4大高危诱因”、“常吃5种水果,女性远离乳腺癌”等报道宣传食疗保健等预防措施。2016年致病因素类报道仍居首位,而治疗方式相关新闻居第2位。“甜食吃多易患乳腺癌”、“‘白骨精’最易得乳腺癌,精神压力大是诱因”等新诱因被发现。“乳腺癌治疗需重视全程管理”、“乳腺癌手术用上‘3D透视技术’”等多种治疗方式应用于临床。2017年乳腺癌预防保健信息仍围绕食用各类果蔬与调节身心健康,治疗方式相关新闻主要介绍精准治疗、药物靶向治疗并提醒患者防止过度治疗。2018年研究进展类新闻居首位,“FDA批准首款针对特定遗传突变的乳腺癌疗法”、

“新研究确认 110 个基因与乳腺癌风险相关”等乳 腺癌防治研究取得突破。整体判断致病因素、预 防保健、治疗方式相关新闻最易成为头条新闻。

表 2 头条新闻主题及内容

主题	频次	主要内容
致病因素	90	烟酒成乳腺癌诱因；高脂饮食易患乳腺癌；日肥胖女性乳腺癌发病率高
预防保健	89	巧吃 7 种天然水果轻松防治乳腺癌；坚持运动是预防乳腺癌；肠道健康降低乳腺癌风险
研究进展	66	美药企 Puma 研发乳腺癌新药获突破；Nature 子刊：乳腺癌精准分层新技术优于常规方法
治疗方式	65	早期乳腺癌患者腋窝放疗优于淋巴结手术；微创乳腺癌根治术成功
特殊患者	55	姚贝娜因乳腺癌复发去世；美 8 岁女孩患乳腺癌；75 岁男子患乳腺癌；3 岁女孩患乳腺癌
疾病特征	45	女性乳腺癌位居榜首；我国乳腺癌发病率每年递增 4%；得了乳腺癌最多还能活多久
诊断检测	41	血检锌含量诊断乳腺癌；尿检可查乳腺癌；早期乳腺癌或不应进行肿瘤标记物检测
病例报道	35	英乳腺癌母亲曾因坚持怀孕推迟治疗；英患乳腺癌晚期女子积极乐观面对病痛感动无数人
公益活动	26	阿联酋“粉红马队”宣传乳腺癌早期防治知识；500 万捐资补助乳腺癌患者
专家辟谣	21	便秘致乳腺癌？缺乏依据；胸小患乳腺癌风险低？缺乏科学依据；长期喝豆浆易得乳腺癌？缺乏依据
行业发展	16	罗氏收购乳腺癌制药公司；新开源投资瑞典企业涉足乳腺癌活检设备
自检自查	11	摸摸腋窝就可以自查乳腺癌；年轻乳腺癌”高发又凶险 自检 + 定期检查可早发现
广告产品	9	广州治疗乳腺癌的最好肿瘤医院；别了乳腺癌！关爱女性健康，中脉美体内衣是认真的
术后防治	5	乳腺癌术后如何防治复发转移；沪医附院成功实施乳腺癌根治术后一期再造手术
保险赔付	4	乳腺癌客户获保险重疾赔付 19 万；客户患乳腺癌新华保险收到理赔材料当天赔付 4 万元

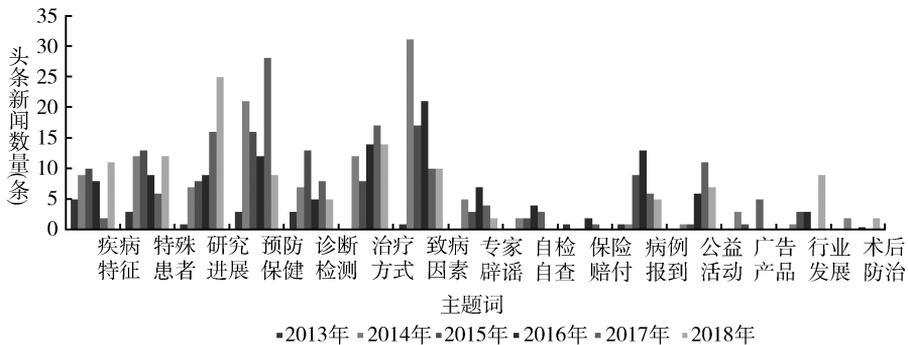


图 6 主题词年度变化

4.2.2 信息源 分类统计头条新闻来源，高频信息源及其性质，见表 3。40.01% 的乳腺癌防治相关头条新闻来自综合类网站，主流媒体均承担乳腺健康宣教责任。网易、新浪等媒体通过转载多源健康信息、发布独家新闻，采用图文结合、吸睛标题等

形式吸引网民关注。和讯网虽主要发布财经类新闻，也多次报道男性患病风险，乳腺增生与乳腺癌的关联性等健康信息。大河网为河南省重点新闻网站，多发布乳腺癌防治建议、食疗预防等信息，也宣传乡镇乳腺癌防治知识科普活动。

表 3 主要信息源及媒体性质

信息源	频次	媒体性质	信息源	频次	媒体性质	信息源	频次	媒体性质
网易新闻	65	综合类	凤凰网	37	综合类	中国新闻网	19	综合类
新浪新闻	44	综合类	新华网	27	综合类	和讯网	15	财经类
人民网	38	综合类	腾讯新闻	24	综合类	大河网	9	地方媒体

5 结语

观察期内网民搜索指数与媒体指数虽出现波动性变化,但近几年均趋于稳定值。随着乳腺癌防治健康信息的普及,公众了解其预防及治疗措施,信息搜索频率在固定范围内波动,媒体的健康知识宣传取得较好成效,信息曝光率降低。公众信息行为与其社会角色相关,网民多关注疾病特征,媒体更注重社会效应。网民搜索主题围绕乳腺癌的早期症状及预后效果,搜索行为受乳腺癌发病率、死亡率及地域、性别、年龄等因素影响。30 岁以下、50 岁以上女性网民及男性网民也需多关注乳腺健康知识。媒体紧密关注医学、药学等领域研究进展,在短时间内大量报道相关主题以促进健康知识大范围传播。全国性、地方性、专业性媒体对乳腺癌防治健康信息的科普宣教力度及信息全面性需进一步加强。

参考文献

- 1 中国互联网信息中心. 第 43 次《中国互联网络发展统计报告》[EB/OL]. [2019-02-28]. http://www.cnnic.net.cn/hlw-fzyj/hlwzxbg/hlwjtjbg/201902/t20190228_70645.htm.
- 2 中国科学技术协会. 中国网民科普需求搜索行为报告

(2017 年度) [EB/OL]. [2019-02-28]. http://www.cast.org.cn/art/2018/3/22/art_341_47213.html.

- 3 龚韩湘, 吴泽埔, 伍宝玲, 等. 我国养老服务模式社会关注度的时空差异分析[J]. 中国卫生政策研究, 2018, 11 (5): 17-21.
- 4 王云娣. 公共文化服务体系的网络关注度研究—以百度指数为例 [J]. 现代情报, 2017, 37 (1): 37-40, 88.
- 5 李清光. 消费者对食品安全信息搜寻行为的特征分析—基于微博指数与百度指数的分析 [J]. 价格理论与实践, 2018 (9): 78-81.
- 6 陈万青, 郑荣寿, 张思维, 等. 2011 年中国恶性肿瘤发病和死亡分析 [J]. 中国肿瘤, 2015, 24 (1): 1-10.
- 7 陈万青, 郑荣寿, 张思维, 等. 2012 年中国恶性肿瘤发病和死亡分析 [J]. 中国肿瘤, 2016, 25 (1): 1-8.
- 8 陈万青, 郑荣寿, 张思维, 等. 2013 年中国恶性肿瘤发病和死亡分析 [J]. 中国肿瘤, 2017, 26 (1): 1-7.
- 9 陈万青, 孙可欣, 郑荣寿, 等. 2014 年中国分地区恶性肿瘤发病和死亡分析 [J]. 中国肿瘤, 2018, 27 (1): 1-14.
- 10 孙可欣, 郑荣寿, 张思维, 等. 2015 年中国分地区恶性肿瘤发病和死亡分析 [J]. 中国肿瘤, 2019, 28 (1): 1-11.
- 11 师金, 梁迪, 李道娟, 等. 全球女性乳腺癌流行情况研究 [J]. 中国肿瘤, 2017, 26 (9): 683-690.

(上接第 5 页)

- 8 Chen Y, Ding S, Xu Z, et al. Blockchain - based Medical Records Secure Storage and Medical Service Framework [J]. J Med Syst, 2018, 43 (1): 5.
- 9 Patel V. A Framework for Secure and Decentralized Sharing of Medical Imaging Data via Blockchain Consensus [J]. Health Informatics J, 2019, 25 (4): 1398-1411.
- 10 高梦婕, 王化群. 基于区块链的可搜索医疗数据共享方案 [J]. 南京邮电大学学报 (自然科学版), 2019, 39 (6): 94-103.
- 11 Wang H, Song Y. Secure Cloud - based EHR System Using Attribute - based Cryptosystem and Blockchain [J]. J Med Syst, 2018, 42 (8): 152.
- 12 甘霖, 杨建华, 路松峰. 基于区块链的电子病历共享技术研究 [J]. 中国数字医学, 2019, 14 (12): 11-13.
- 13 Narayanaswami C, Raj Nooyi, Raghavan SG, et al. Blockchain Anchored Supply Chain Automation [J]. IBM J Res Dev, 2019 (99): 1.
- 14 Nugent T, Upton D, Cimpoesu M. Improving Data Trans-

parency in Clinical Trials Using Blockchain Smart Contracts [J]. F1000Res, 2016 (5): 2541.

- 15 Benchoufi M, Porcher R, Ravaut P. Blockchain Protocols in Clinical Trials: transparency and traceability of consent [J]. F1000Res, 2017 (6): 66.
- 16 Sylim P, Liu F, Marcelo A, et al. Blockchain Technology for Detecting Falsified and Substandard Drugs in Distribution: pharmaceutical supply chain intervention [J]. JMIR Res Protoc, 2018, 7 (9): e10163.
- 17 Jo B W, Khan R, Lee Y S. Hybrid Blockchain and Internet - of - Things Network for Underground Structure Health Monitoring [J]. Sensors (Basel), 2018, 18 (12): 4268.
- 18 Khan M A, Salah K. IoT Security: review, blockchain solutions, and open challenges [J]. Future Generation Computer Systems, 2018 (82): 395-411.
- 19 Bogoeva A. Blockchain in Healthcare: opportunities and challenges [D]. San Antonio: The University of Texas at San Antonio, 2018.