参考文献

[1] Bellingrath S, Weigl T, Kudielka BM. Cortisol dysregulation in school teachers in relation to burnout, vital exhaustion, and effort-reward-imbalance[J]. Biol Psychol, 2008, 78(1): 104 – 113.

[2] 黄晓艳,颜鸿飞,王秀珍,等. 精神科医生职业紧张与工作绩效关系[J]. 中国公共卫生, 2015, 31(8): 994 – 996.

[3] Stojek M, Fischer S, Collins B. Thinness and restricting expectancies mediate the influence of ethnic identity on bulimic symptoms[J]. Pers Indiv Differ, 2010, 49(2): 102 – 106.

[4] 张金美,丁卫星,陈方斌,等. 军人职业倦怠量表的信效度检验[J]. 中华行为医学与脑科学杂志, 2010, 19(11): 1034 – 1036. DOI: 10.3760/cma.j.issn.1674 – 6554.2010.11.028.

[5] 张理义,梅贵森,任忠文,等. 少数民族军人心身健康状况及其影响因素分析[J]. 解放军医学杂志, 2013, 38(2): 156 – 160.

[6] Wissink IB, Dekovci M, Yagmur S. Ethnic identity, externalizing problem behaviour and the mediating role of self-esteem among Dutch, Turkish-Dutch and Moroccan-Dutch adolescents[J]. J Adolesc, 2008, 31(2): 223 – 240.

[7] Reininghaus U, Craig TKJ, Fisher HL, et al. Ethnic identity, perceptions of disadvantage, and psychosis findings from the SOP study[J]. Schizophr Res, 2010, 124: 43 – 48.

[8] Gramer M, Reitbauer C. The influence of social support on cardiovascular responses during stressor anticipation and active coping[J]. Biol Psychol, 2010, 85(2): 268 – 274.

[9] 李明军. 中小学教师工作家庭冲突、职业倦怠与生活满意度的关系[J]. 中国健康教育, 2015, 31(9): 830 – 833.

[10] 曲海英,刘修英. 三级甲等医院护士成就动机和职业倦怠的关系[J]. 中国公共卫生, 2015, 31(6): 807 – 810.

[11] 苏宗义,丁魁,顾焱,等. 南北疆军人职业倦怠比较研究[J]. 中华行为医学与脑科学杂志, 2014, 23(5): 451 – 453. DOI: 10.3760/cma.j.issn.1674 – 6554.2014.05.021.

[12] 李玉莲,吕筠,沈艳辉,等. 基层疾病预防控制机构人员职业倦怠状况及影响因素分析[J]. 中国公共卫生, 2014, 30(12): 1545 – 1549.

[13] 童辉杰,刘轩. 大学生村官工作适应与职业倦怠关系[J]. 中国公共卫生, 2014, 30(3): 257 – 259.

[14] 宋之杰,姚晶晶,石蕊,等. 消防官兵心理健康与社会支持、应对方式关系研究[J]. 中国职业医学, 2013, 40(6): 533 – 536. DOI: 10.11763/j.issn.2095 – 2619.2013.06.011.

收稿日期: 2015 – 12 – 24　　(郑新编辑　张翠校对)

中国公共卫生 CPH

· 调查报告与分析 ·

特定关键词及百度指数与流感病毒活动相关性分析*

董晓春[1], 李琳[1], 徐文体[1], 张颖[1], 张之伦[1], 赵卓[2]

摘　要: 目的　分析特定关键词的搜索引擎数据与流感病毒活动的相关性,探讨百度指数在流感监测中的应用。**方法**　根据对流感确诊病例调查选择最优关键词,通过百度搜索得到各关键词的百度指数。并采用 Spearman 相关分析和数据移动对天津市 2014—2015 年流感流行季百度指数与哨点监测数据进行相关分析。**结果**　流感确诊病例中 65.04% (173/266) 的人在出现症状后通过网络搜索引擎寻求过帮助。"感冒"作为关键词占比最高为 67.72% (107/158)。关键词"发烧"的百度指数和阳性率的相关系数最高 ($r = 0.808$, $P < 0.05$), 阳性率向前移动 1 周,与"发烧"百度指数的相关性增加 ($r = 0.827$, $P < 0.05$)。**结论**　关键词"发烧"的百度指数能较好的反映和更早的预测流感病毒的活动。

关键词: 百度指数; 搜索引擎; 流感病毒活动

中图分类号: R 181.2　**文献标志码:** A　**文章编号:** 1001-0580(2016)11-1543-04　DOI: 10.11847/zgggws2016-32-11-25

Correlation between Baidu index of specific keywords and influenaz epidemics

DONG Xiao-chun*, LI Lin, XU Wen-ti, et al (*Department of Infectious Disease, Tianjin Municipal Center for Disease Control and Prevention, Tianjin* 300011, China)

Abstract: Objective　To analyze the correlation between Baidu index of specific keywords and influenza epidemics and to explore the utilization of the index in influenza surveillance. **Methods**　Data on reported influenza like illness (ILI) cases and weekly positive rate of influenza virus isolation among the cases from October 2014 to March 2015 were collected form influenza surveillance sentinel hospitals in Tianjin city and 266 laboratory test confirmed influenza cases were surveyed with a questionnaire. Search queries related to influenza epidemics were gathered from Baidu database for the calculation of Baidu keyword index. Spearman correlation was used to analyze the relationship between influenza epidemics and Baidu keyword index. **Results**　Among the confirmed influenza cases, 65.04% performed internet searches about their early symptoms using search engine and "cold" was a keyword with the highest query frequency (67.72%) used by the cases. There was a highly close correlation between Baidu index for "fever" and the weekly positive rate of

* **基金项目:** 国家科技重大专项(2012ZX10004 – 209); 天津市卫生局科技基金(2013KZ049)

作者单位: 1. 天津市疾病预防控制中心传染病科, 天津 300011; 2. 辽宁省疾病预防控制中心

作者简介: 董晓春(1981 –), 男, 山西忻州人, 主管医师, 硕士, 主要从事传染病预防控制工作。

通讯作者: 赵卓, E-mail: zhaozhuo699@163.com

数字出版日期: 2016 – 4 – 6 7:39

数字出版网址: http://www.cnki.net/kcms/detail/21.1234.R.20160406.0739.006.html

influenza virus isolation ($r = 0.808$, $P < 0.05$) and the correlation was more close when a one-week lag Baidu index for "fever" was adopted in the analysis ($r = 0.827$, $P < 0.05$). **Conclusion**　Baidu index for "fever" could be considered as a earlier signal for incidence of influenza epidemic.

Key words: Baidu index; search engine; influenza epidemic

流行性感冒简称流感,是由流感病毒引起的一种急性呼吸道传染性疾病。流感的局部暴发或流行几乎每年均会发生[1]。常规流感监测包括病原学和流感样病例监测,然而这两类数据均来自于前往哨点医疗机构就诊病例信息。而现在越来越多的人通过搜索引擎寻求健康和医疗信息,网络搜索词量的动态变化一定程度上反映该地区相关疾病流行情况和人群中发病及求医的信息。因此一种新的监测方法—应用网络搜索词信息辅助传染病监测为加强监测能力提供契机[2]。本研究分析天津市 2014—2015 年流感流行季节与流感相关的特定关键词的百度指数与流感病毒活动的相关性,现将结果报告如下。

1　资料与方法

1.1　流感哨点监测

1.1.1　流感样病例监测　流感样病例(influenza like illness, ILI)定义为发热(体温≥38 ℃),伴咳嗽或咽痛之一者。天津市 9 家国家级流感监测哨点医院全年进行流感样病例监测。哨点医院监测诊室的医务人员,按照 ILI 的定义,每天按科室登记各年龄组的流感样病例数和门急诊病例就诊总数,由哨点医院主管科室每日收集、汇总后,将本院各监测诊室数据录入到"中国流感监测信息系统"。本次研究选取国家级流感监测哨点医院 2014—2015 年度流感流行季节(2014 年 10 月—2015 年 3 月)流感样病例监测数据。共调查流感实验室诊断病例 266 例。

1.1.2　流感病原学监测　流感监测哨点监测医院每周采集 ILI 标本,数量要求为每周至少采集 20 份。全部标本送流感网络实验室以实时反转录酶-聚合酶链锁反应(real-time reverse transcription-polymerase chain reaction, RT-PCR)方法进行流感病毒核酸检测,并每周通过"中国流感监测信息系统"报告检测结果。

1.2　百度指数监测

1.2.1　关键词筛选　通过对流感实验室诊断病例进行问卷调查,问卷自行设计并知情同意。问题涉及:"在出现症状后是否通过网络搜索引擎寻求过帮助?"如果有,在提供的"流感"、"感冒"、"禽流感"、"肺炎"、"H7N9","发烧"、"咳嗽"、"嗓子疼"、"头疼"、"新型流感"10 个关键词中,倾向于选择哪个词或哪几个作为检索关键词,如以上未包括请写出其他关键词。筛选超过 50% 的前 5 位关键词进行百度指数的采集。

1.2.2　百度指数收集　通过设置地点参数"天津",时间参数"2014 年 10 月—2015 年 3 月",用筛选的到的关键词从百度指数数据库(http://index.Baidu.com/)中收集各关键词的每日百度指数。

1.2.3　以"周"为单位建立流感病毒活动(周阳性率)和相关"百度指数"以及流感样病例百分比的时间序列:

$$\text{流感病毒周阳性率} = \frac{\text{实验室检周流感阳性数}}{\text{每周流感样病例采样数}} \times 100\%$$

$$\text{流感样病例百分比(ILI\%)} = \frac{\text{哨点门急诊周 ILI 人数}}{\text{哨点门急诊周就诊总人数}} \times 100\%$$

周百度指数为 1 周 7 日百度指数的合计值。

1.3　统计分析　采用 EXCEL 2007 建立数据库,对数据进行描述性统计,相关分析采用 Spearman 相关计算相关系数,$P < 0.05$ 有统计学意义。全部统计分析由 SPSS 19.0 统计软件完成。

2　结　果

2.1　关键词筛选结果　共调查流感实验室诊断病例 266 例,其中 65.04%(173/266)在出现症状后通过网络搜索引擎寻求过帮助,173 例中有 158 例回答了关键词,应答率为 91.33%。结果显示,调查表中列举的 10 个搜索关键词,排在前 5 位的分别为"感冒"67.72%(107/158)、"发烧"62.03%(98/158)、"流感"52.53%(83/158)、"咳嗽"51.90%(82/158)、"嗓子疼"19.38%(48/158);其余 5 个关键词占比均 <10%。自行填写的关键词中,"退热/退烧"占比最高,为 5.06%(8/158)。综合占比排名,将"感冒""发烧""流感""咳嗽"作为关键词进行百度指数的查询。"嗓子疼"占比未超 50%,但作为流感样病例定义中的其中 1 个症状,也作为本次研究查询的关键词。

2.2　哨点监测与百度指数数据变化趋势　2014 年 10 月—2015 年 3 月期间,流感病毒阳性率平均为 23.27%(1 124/4 830),其中 2014 年 50 周(1450 周)最高,为 48.69%(93/191);ILI% 平均为 5.12%(113 253/2 211 010),在 1 452 周达到峰值 7.52%(8 532/113 392)。百度指数方面:"发烧"平均为 222,最高为 369(1 451 周);"咳嗽"平均为 224,最高为 285(1 452 周);"感冒"平均为 261,最高为 379

(1 451 周);“流感”平均为 146,最高为 228(1 451 周);“嗓子疼”平均为 154,最高为 177(1 451 周)。除百度指数“嗓子疼”外,其他指数以及 ILI% 和病毒阳性率均呈现明显的单一高峰。

2.3　百度指数及 ILI% 与病毒阳性率的关系(图 1)　各指标与病毒阳性率的 Spearman 相关分析结果显示,ILI% 与病毒阳性率的相关有统计学意义(r = 0.640, P < 0.05);百度指数中,5 个关键词与病毒阳性率均有相关关系,相关系数 r 从大到小分别为“发烧”0.808,“感冒”0.666,“流感”0.624,“咳嗽”0.467,“嗓子疼”0.462(均 P < 0.05)。

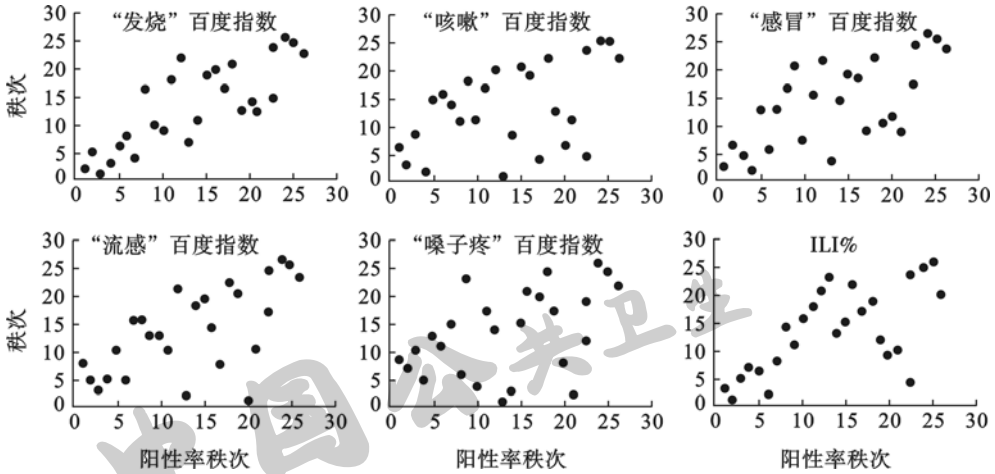

图 1　百度指数及 ILI% 与病毒阳性率的散点图

2.4　病毒阳性率移动后与各指标关系(表 1)　病毒阳性率向前移动 1 周,相关系数增加的指标:ILI% 从 0.640 增加到 0.688;百度指数中,“发烧”从 0.808 增加到 0.827,“咳嗽”从 0.467 增加到 0.468,“感冒”从 0.666 增加到 0.667。而“流感”和“嗓子疼”2 个指数的相关系数有所降低。以上相关均有统计学意义(P < 0.05)。病毒阳性率向后移动 1 周,相关系数全部降低。且“咳嗽”与病毒阳性率变为无相关性(P > 0.05)。

表 1　百度指数及 ILI% 与病毒阳性率的相关系数(r)

病毒阳性率	百度指数					ILI%
	发烧	咳嗽	感冒	流感	嗓子疼	
向前移动 1 周	0.827[a]	0.468[a]	0.667[a]	0.582[a]	0.394[a]	0.688[a]
不移动	0.808[a]	0.467[a]	0.666[a]	0.624[a]	0.462[a]	0.640[a]
向后移动 1 周	0.711[a]	0.381	0.625[a]	0.556[a]	0.458[a]	0.538[a]

注:a P < 0.05。

3　讨　论

随着全球所面临的生物恐怖袭击和传染病爆发流行威胁的增多,直接促使了症状监测在理论与实践方面取得了很大进展,技术手段趋于成熟。与传统公共卫生监测不同,症状监测收集各种与健康事件相关的数据,尤其是指一种传染病在实验室确诊之前,患者可能表现出的行为方式、症状、体征或实验室结果的异常,依据多种数据源对某种公共卫生事件进行监测,以便对疾病暴发进行早期探查、预警及快速反应[3]。已有多种症状监测方法用于实时监测和预测流感传播。这其中就包括基于搜索引擎和互联网的流感相关关键词的检索量监测。互联网的发展已经极大地改变了人们搜索医疗信息的方式,越来越多的信息可以通过搜索引擎在互联网上查询到,特别是一些关于疾病的相关信息,这为互联网搜索监测提供了可行性基础。据中国互联网络信息中心发布的信息,截至 2015 年 6 月,中国网民规模达 6.68 亿[4]。既往就有国内外研究将谷歌流感趋势(Google flu trend)的应用于流感的症状监测。但在中国,谷歌的搜索引擎由于使用率较低,基于谷歌的症状监测并不适用。百度在中国的渗透率和网民常用率上高居首位,过去半年使用过百度的搜索网民比例(渗透率)达 97.9%,在网民常用(首选)的搜索引擎中,百度占据了 84.5% 以上的份额,网民选择较为集中[5]。百度指数是以百度网页搜索和百度新闻搜索为基础的免费海量数据分析服务,用以反映不同关键词在过去一段时间里的“用户关注度”和“媒体关注度”。发现、共享和挖掘互联网上有价值的信息和资讯,直接、客观地反映网民的兴

趣和需求。本研究就基于特定关键词的百度指数探讨其在监测流感病毒活动中的应用。

本研究采用对实验室确诊病例进行问卷的方法筛选出最可能代表居民病后搜索行为的关键词,从方法上避免了研究者的主观因素影响。从调查结果来看,65.04% 的人在出现症状后通过网络搜索引擎寻求过帮助,这一比例高于美国成年人通过互联网获取医学信息的比例 47.8%[6],这得益于天津市网络的建设和普及,《天津市互联网发展状况报告》的数据显示,截止 2014 年底,天津市网民达到 904 万人;网民普及率 61.4%,高出全国平均水 13.5 个百分点[7]。根据筛选规则,"感冒""发烧""流感""咳嗽"被筛选为本研究的关键词,这一结果与韩国的一项调查结果基本一致[8]。同时为避免漏选,增加"嗓子疼"作为补充关键词。为更准确的探讨百度指数和流感病毒活动的关系,尽量减少其他呼吸道病原引起搜索波动的混杂,特选择流感流行季作为研究的时间段。从研究结果来看,所有 5 个关键词的百度指数与流感病毒活动均有相关性,其中"发烧"的相关性最高,其次为"感冒"。两者与流感病毒活动的相关性均高于 ILI% 数据。这一结果与广东的一项基于谷歌流感趋势的研究有所不同。可能与其研究的时间段以及地域不同有关[9]。发烧症状是流感发病急性期主要的临床表现,也是流感样病例的诊断的必备条件之一,虽然发烧并不是流感的特异性症状,但发烧症状的易识别和易受人关注的特点使其成为反映流感病毒活动的最佳指标。"感冒"作为医生用来描述急性上呼吸道症候群的传统术语,早已为大众熟知,在无法区别感冒及流感的概念时,通常大众会以感冒作为一系列广泛的呼吸道感染的代名词,进而在互联网寻求医学帮助,从本次调查结果"感冒"居搜索关键词首位就印证了这一点。在流感流行季,流感病毒的活动增大了其在感冒这一症候群中的比例,可能是造成"感冒"这一关键词与流感病毒活动有较高相关性的原因。

从流感病毒阳性率移动结果来看,阳性率向前移动 1 周,"发烧"、"感冒"、"咳嗽"的百度指数与其相关系数均有所增加,表明这 3 个关键词在监测到流感病毒活动前 1 周就提前出现变化。综合相关系数判断,"发烧"这一关键词变化能够更早和更敏感的反映流感病毒阳性率的变化。由于目前网络资源的丰富和查询的便利,个人在出现流感的早期症状时可能首先将他们的症状通过互联网搜索寻求帮助,继而会在这之后的几天内到医院就诊[10],因此,搜索行为较就诊数据更早的体现流感病毒的活动的变化。

综合以上研究结果,在流感流行期间,网络搜索信息较常规监测更加灵敏,且无需依赖哨点监测的支持,且搜索信息为免费获取,成本低廉,在传统监测的基础上,基于搜索引擎的百度指数监测可以成为其有效的补充。

参考文献

[1] World Health Organization. Influenza (seasonal) fact sheet No. 211, WHO website[EB/OL]. http://www.who.int/mediacentre/factsheets/fs211/en/, Accessed March 2014.

[2] Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet[J]. CMAJ, 2009, 180(8): 829 - 831.

[3] 龙璐,严薇荣,许奕华,等. 症状监测系统预测预警模型研究进展[J]. 中国公共卫生, 2012, 28(5): 704 - 706.

[4] 中国互联网络信息中心. 第 36 次中国互联网络发展状况调查统计报告[EB/OL]. [2015 - 07 - 22], http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201507/P020150723549500667087.pdf.

[5] 中国互联网络信息中心. 2014 年中国网民搜索行为研究报告[EB/OL]. [2015 - 03 - 16], http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/201410/P020150104459023035664.pdf.

[6] Lion S, Jamie B, Benjamin G, et al. Seeking health information and support online: does it differ as a function of engagement in risky health behaviors? Evidence from the health information national trends survey[J]. Journal of Medical Internet Research, 2014, 16(11): e235.

[7] 中国国家互联网络信息办公室. 天津市互联网发展状况报告[EB/OL]. [2015 - 05 - 22], http://www.cac.gov.cn/2015 - 05/22/c_1115369561.htm.

[8] Cho S, Sohn CH, Jo MW, et al. Correlation between national influenza surveillance data and Google trends in South Korea[J]. PLoS One, 2013, 8(12): e81422.

[9] Kang M, Zhong H, He J, et al. Using Google trends for influenza surveillance in South China[J]. PLoS One, 2013, 8(1): e55205.

[10] Thompson LH, Malik MT, Gumel A, et al. Emergency department and 'Google flu trends' data as syndromic surveillance indicators for seasonal influenza[J]. Epidemiology and Infection, 2014, 142: 2397 - 2405.

收稿日期: 2015-10-28　　(吴少慧编辑　张翠校对)

中国公共卫生 CPH