Public Health and Epidemiology Informatics

R. Thiébaut[1,2,3], F. Thiessard[1,2], Section Editors for the IMIA Yearbook Section on Public Health and Epidemiology Informatics

[1] Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219, Bordeaux, France
[2] Centre Hospitalier Universitaire de Bordeaux, Service d'Information Médicale, Bordeaux, France
[3] Inria, SISTM, Talence, France

Summary

Objectives: To summarize current research in the field of Public Health and Epidemiology Informatics.
Methods: The complete 2016 literature concerning public health and epidemiology informatics has been searched in PubMed and Web of Science, and the returned references were reviewed by the two section editors to select 14 candidate best papers. These papers were then peer-reviewed by external reviewers to allow the editorial team an enlightened selection of the best papers.
Results: Among the 829 references retrieved from PubMed and Web of Science, three were finally selected as best papers. The first one compares Google, Twitter, and Wikipedia as tools for Influenza surveillance. The second paper presents a Geographic Knowledge-Based Model for mapping suitable areas for Rift Valley fever transmission in Eastern Africa. The last paper evaluates the factors associated with the visit of Facebook pages devoted to Public Health Communication.
Conclusions: Surveillance is still a productive topic in public health informatics but other very important topics in public health are appearing.

Keywords

Public health; epidemiology; medical informatics; International Medical Informatics Association; health information systems

Yearb Med Inform 2017:248-51
http://dx.doi.org/10.15265/IY-2017-036
Published online August 18, 2017

Introduction

Public Health Informatics is the systematic application of information and computer sciences to public health practice, research, and learning, as it has been quoted in the Medical Subject Heading information since 2013. A new section on this specific topic has been integrated in the IMIA Yearbook in 2015. An overview of the area covered by Public Health Informatics is presented in the Dixon et al. paper [1]. In 2016, the term Precision Public Health started to appear in several publications [2, 3]. In parallel to precision medicine, it is concerned with providing the right intervention to the right population at the right time [2]. The idea is to take advantage of the development of technologies including health information technology to improve the assessment of population health and prevention interventions and policies. Surveillance of epidemics and community health issues are obvious examples that are pinpointed in publications [2, 3]. New technologies and big data should help accelerate the detection of epidemics in a timely and accurate manner by accessing laboratory, satellite, and phone data, tracking population movements, and integrating all data for making more precise estimations. Modelling the risk of epidemics in well-defined areas could help in targeting interventions for preventing epidemics. Hence, although public health informatics covers a large spectrum of applications, the surveillance of epidemics using recently available web-based and other tools, that could be referred to as precision epidemiology or digital epidemiology, constitutes a recurrent topic in the literature. In addition, numerous papers have been published and present methods to optimize and analyze internet data for various infectious agents, as it is done with Google flu[1]. This "classical epidemiology of infectious disease" using new digital tools should eventually be useful for public agencies and for the surveillance of other diseases. But public health informatics is covering many other areas of research including communication. Using newly available tools and especially web-based ones should be of benefit to the public health informatics community.

Paper Selection

A comprehensive literature search was performed using two bibliographic databases, Pubmed/Medline (from NCBI, National Center for Biotechnology Information), and Web of Science® (from Thomson Reuters). The papers had to be journal articles, excluding all other kinds of papers (such as comments, letters, case reports, etc.), written in English, and having an abstract. The following keywords were selected for the query: public health informatics or at least one of "public health, epidemiology, disease outbreaks, registries, epidemiologic study characteristics, epidemiological monitoring, population surveillance, public health surveillance, sentinel surveillance, public health practice, organizational policy, planning techniques", and at least one of "medical records systems, computerized, computing methodologies, signal processing, computer-assisted, mathematical computing, computer simulation, expert systems, fuzzy logic, knowledge bases, neural

[1] https://www.google.org/flutrends/about/

networks (computer), medical informatics, medical informatics computing, medical informatics applications, decision support techniques, community networks, databases as Topic, information dissemination, health information systems" or "techniques such as Fourier, cyclic analysis, neural networks, data sources as Internet, social network, knowledge bases, computerized medical record system, and telemedicine".

The search was targeted at public health and epidemiology papers that involve computer science or the massive amount of web-generated data. References addressing topics of other sections of the Yearbook, such as those related to interoperability between data providers or clinical research were excluded from our search. The study was performed at the beginning of January 2017, covering the year 2016. A total of 807 references were returned.

Articles were separately reviewed by the two section editors, and were first classified into three categories: "keep", "discard", or "leave pending". Then, the two lists of references were merged, yielding 73 references that were retained by at least one reviewer or classified as "leave pending" by both of them. The two section editors jointly reviewed the 73 references and drafted a consensual list of 14 candidate best papers. All pre-selected 14 papers were then peer-reviewed by editors and external reviewers (at least four reviewers per paper). Three papers were finally selected as best papers (Table 1). A content summary of these selected papers can be found in the appendix of this synopsis. Lamy, et al., [4] describe the entire selection process.

Outlook and Conclusion

A substantial number of short-listed papers were about digital surveillance of infectious diseases. Several compared the sources of information among Google, Twitter, and Wikipedia [5-9] for various infectious diseases such as Middle East Respiratory Syndrome Coronavirus (MERS-CoV) in Korea, bubonic plague outbreak in Madagascar, chicken pox caused by varicella zoster virus (VZV), and Influenza in United States. Of note, one work presented an open-source system gathering tweets on symptoms associated with influenza-like illness (ILI) [8], and another used Twitter for avian influenza risk surveillance [9]. The work performed by Sharpe *et al.*, is very well done and particularly interesting because it adds an unconventional source of information: the accesses to Wikipedia pages on Influenza [5]. In addition to the objective of detecting outbreaks as early as possible, the same approaches could be applied to look at the impact of vaccination programs [10]. Besides web-based data, information extracted from cloud-based electronic health record (EHR) databases can also be used for real-time surveillance of influenza-like illnesses [11]. In the past, it was not possible to use medical records for tracking epidemics because of the time lag due to the availability of data. Today, the easy access to EHRs makes their use for real time surveillance possible.

Another example of digital surveillance is active surveillance using short message service (SMS), or text messages, as described by Caceres *et al.*, [12] in the context of the recent Ebola epidemics for daily reporting of zero cases. They have shown that such surveillance was feasible and may be rapidly implemented even in low resource countries.

Surveillance of infectious diseases is also performed through geographic information systems (GIS). Tran et *al.*, [13] present an adaptation of a geographic knowledge-based method [14] to identify areas for Rift Valley fever transmission in Eastern Africa. Allen *et al.*, found a statistically significant correlation between influenza outbreaks using the social media platform Twitter and techniques from GIS for the thirty most populated cities in the United States during the 2013–2014 influenza season, compared with national, regional, and local influenza outbreak reports [15]. A visual analytics GIS-based decision support system for early infectious diseases outbreak detection was applied in Pakistan, using real-time streaming data from emergency departments [16].

GISs are also used for chronic diseases as Laranjo *et al.*, demonstrate for type 2 diabetes [17]. Akil *et al.*, used GISs to show that geographic location besides socioeconomic status may contribute to the high rates of Salmonella in Mississippi [18].

Besides surveillance, a paper about communication has been selected as one of this year best paper [19]. This study aims at identifying the features of Facebook posts that are associated with higher user engagement on Australian public health organizations' Facebook pages.

Table 1 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2017 in the section 'Public Health and Epidemiology Informatics'. The articles are listed in alphabetical order of the first author's surname.

References	Topic
▪ Kite J, Foley BC, Grunseit AC, Freeman B. Please Like Me: Facebook and Public Health Communication. PLoS One 2016;11(9).	Prevention
▪ Sharpe JD, Hopkins RS, Cook RL, Striley CW. Evaluating Google, Twitter, and Wikipedia as Tools for Influenza Surveillance Using Bayesian Change Point Analysis: A Comparative Analysis. JMIR Public Health Surveill 2016 20;2(2).	Surveillance
▪ Tran A, Trevennec C, Lutwama J, Sserugga J, Gély M, Pittiglio C, Pinto J, Chevalier V. Development and Assessment of a Geographic Knowledge-Based Model for Mapping Suitable Areas for Rift Valley Fever Transmission in Eastern Africa. PLoS Negl Trop Dis 2016;10(9).	Surveillance

Acknowledgements

We would like to thank all external reviewers for their participation in the selection process of the best papers for the Public Health and Epidemiology Informatics section of the IMIA Yearbook.

References

1. Dixon BE, Kharrazi H, Lehmann HP. Public Health and Epidemiology Informatics: Recent Research and Trends in the United States. Yearb Med Inform 2015;10(1):199206.
2. Khoury MJ, Iademarco MF, Riley WT. Precision Public Health for the Era of Precision Medicine. Am J Prev Med 2016;50(3):398401.
3. Dowell SF, Blazes D, Desmond-Hellmann S. Four steps to precision public health. Nature News 2016;540(7632):189.
4. Lamy J-B, Séroussi B, Griffon N, Kerdelhué G, Jaulent M-C, Bouaud J. Toward a formalization of the process to select IMIA Yearbook best papers. Methods Inf Med 2015;54(2):13544.
5. Sharpe JD, Hopkins RS, Cook RL, Striley CW. Evaluating Google, Twitter, and Wikipedia as Tools for Influenza Surveillance Using Bayesian

Change Point Analysis: A Comparative Analysis. JMIR Public Health Surveill 2016;2(2):e161.
6. Shin S-Y, Seo D-W, An J, Kwak H, Kim S-H, Gwack J, et al. High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. Sci Rep 2016;6:32920.
7. Da'ar OB, Yunus F, Md Hossain N, Househ M. Impact of Twitter intensity, time, and location on message lapse of bluebird's pursuit of fleas in Madagascar. J Infect Public Health 2017;10(4):396-402.
8. Chorianopoulos K, Talvis K. Flutrack.org: Open-source and linked data for epidemiology. Health Informatics J 2016;22(4):96274.
9. Robertson C, Yee L. Avian Influenza Risk Surveillance in North America with Online Media. PLoS One 2016;11(11):e0165688.
10. Bakker KM, Martinez-Bakker ME, Helm B, Stevenson TJ. Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. PNAS 2016;113(24):668994.
11. Santillana M, Nguyen AT, Louie T, Zink A, Gray J, Sung I, et al. Cloud-based Electronic Health Records for Real-time, Region-specific Influenza Surveillance. Sci Rep 2016;6:25732.
12. Cáceres VM, Cardoso P, Sidibe S, Lambert S, Lopez A, Pedalino B, et al. Daily zero-reporting for suspect Ebola using short message service (SMS) in Guinea-Bissau. Public Health 2016;138:6973.
13. Tran A, Trevennec C, Lutwama J, Sserugga J, Gély M, Pittiglio C, et al. Development and Assessment of a Geographic Knowledge-Based Model for Mapping Suitable Areas for Rift Valley Fever Transmission in Eastern Africa. PLoS Negl Trop Dis 2016;10(9):e0004999.
14. Malczewski J. GIS-based multicriteria decision analysis: a survey of the literature. Int J Geogr Inf Sci 2006;20(7):703-26.
15. Allen C, Tsou M-H, Aslam A, Nagel A, Gawron J-M. Applying GIS and Machine Learning Methods to Twitter Data for Multiscale Surveillance of Influenza. PLoS One 2016;11(7):e0157734.
16. Ali MA, Ahsan Z, Amin M, Latif S, Ayyaz A, Ayyaz MN. ID-Viewer: a visual analytics architecture for infectious diseases surveillance and response management in Pakistan. Public Health 2016;134:7285.
17. Laranjo L, Rodrigues D, Pereira AM, Ribeiro RT, Boavida JM. Use of Electronic Health Records and Geographic Information Systems in Public Health Surveillance of Type 2 Diabetes: A Feasibility Study. JMIR Public Health Surveill 2016;2(1):e12.
18. Akil L, Ahmad HA. Salmonella infections modelling in Mississippi using neural network and geographical information system (GIS). BMJ Open 2016;6(3):e009255.
19. Kite J, Foley BC, Grunseit AC, Freeman B. Please Like Me: Facebook and Public Health Communication. PLoS One 2016;11(9):e0162765.

Correspondence to:
Rodolphe Thiébaut
Inserm U1219, ISPED, Univ. Bordeaux
146 rue Leo Saignat
33076 Bordeaux cedex, France
Tel: +33 5 57 57 45 21
Fax: +33 5 56 24 00 81
E-mail: rodolphe.thiebaut@u-bordeaux.fr

Appendix: Content Summaries of Selected Best Papers for the 2017 IMIA Yearbook, Section 'Public Health and Epidemiology Informatics'

Kite J, Foley BC, Grunseit AC, Freeman B

Please Like Me: Facebook and Public Health Communication

PLoS One 2016;11(9)

This study aimed at reviewing the use of Facebook by Australian public health organisations to identify features of posting activity that are associated with user engagement, which authors define as likes, shares, or comments. Authors selected 20 eligible pages relevant to selected public health issues through a systematic search and coded 360-days of posts for each page. The health issues were: smoking, healthy diet, physical activity/sedentariness, overweight/obesity, alcohol, sexual health, illicit drug use, skin cancer, aboriginal health. Posts were coded by: post type (photo, text only, game, poll/quiz, app, link, event, or video), communication technique employed (informative, call-to-action, instructive, positive emotive appeal, fear appeal, testimonial, humor), and use of marketing elements (e.g., branding, use of mascots, etc.). Negative binomial regressions were used to assess associations between post characteristics (post type, communication technique, and marketing elements as categorical independent variables), and user engagement (respectively, number of likes, shares, and comments as the outcome variables). The results showed that video posts produced the greatest amount of user engagement, although an analysis of a subset of the data suggested that this might be a reflection of the Facebook algorithm, which governs what is and is not shown in user newsfeeds and appears to prefer videos over other post types. Posts that featured a positive emotional appeal or provided factual information attracted higher levels of user engagement, while conventional marketing elements, such as sponsorships and the use of persons of authority, generally discouraged user engagement, with the exception of posts that included a celebrity or a sportsperson. Further research could assist in understanding whether engagement with public health-related pages on Facebook actually leads to the achievement of public health goals. This study has shown that in order to increase the chances of achieving public health goals, content providers must encourage engagement and adapt to the Facebook algorithm in order to maximize message exposure, while also ensuring that the content is of high quality.

Sharpe JD, Hopkins RS, Cook RL, Striley CW

Evaluating Google, Twitter, and Wikipedia as Tools for Influenza Surveillance Using Bayesian Change Point Analysis: A Comparative Analysis

JMIR Public Health Surveill 2016 20;2(2)

Traditional influenza surveillance relies on the reports provided by health care providers of influenza-like illness (ILI) syndromes. It primarily captures individuals who seek medical care and misses those who do not interact with the health care system, and this surveillance method is limited by relatively dated technology and by delays of up to one to two weeks between the occurrence of the illness event and the dissemination of surveillance information. Syndromic surveillance includes the use of novel data sources such as emergency department records and prescription sales to enhance traditional surveillance systems. Recently, nontraditional data sources, particularly Web-based, have been applied to public health surveillance, as there is a growing number of people who search, post, and tweet about their illnesses before seeking medical care. This so coined 'digital epidemiology' can be less expensive, timelier, and can expand detection by increasing the range of health events that can be detected. Existing research has shown some promise of using data from Google, Twitter, and Wikipedia to complement traditional surveillance for ILI, but none compared the three of them. The objective of this study is to comparatively analyze Google Flu Trends , Twitter, and Wikipedia by examining which best corresponds with Centers for Disease Control and Prevention

(CDC) ILI data. It was hypothesized that Wikipedia will best correspond with CDC ILI data as a previous research found it to be least influenced by high media coverage as compared with Google and Twitter. Publicly available, deidentified data were collected from the CDC, Google Flu Trends, HealthTweets, and Wikipedia for the 2012-2015 influenza seasons. Bayesian change point analysis was used to detect seasonal changes, or change points, in each of the data sources. Change points in Google, Twitter, and Wikipedia that occurred during the exact week, the preceding week, or the week after the CDC's change points were compared with the CDC data as the gold standard. All analyses were conducted using the R package "bcp" version 4.0.0 in RStudio. In addition, sensitivity and positive predictive values (PPV) were calculated for Google Flu Trends, Twitter, and Wikipedia. During the 2012-2015 influenza seasons, a high sensitivity of 92% and a PPV of 85% were found for Google Flu Trends. A low sensitivity of 50% and a low PPV of 43% were found for Twitter. Wikipedia had the lowest sensitivity of 33% and lowest PPV of 40%. Limitations: 1) Bayesian change point analysis assumes time series data are distributed normally, which may not be the case with public health surveillance data, 2) for the analysis of Wikipedia views, only the "Influenza" article was used for analysis, excluding other articles on influenza medications and influenza strains. The authors assumed that all the views of the English-language Wikipedia "Influenza" article were done by US users when some may have come from users in other English-speaking countries where the influenza season is very different, 3) the Google Flu Trends data were fitted to match CDC data, 4) data duplication could be an issue with each data source used in this study, 5) Internet users are younger than the general U.S. population. Of the three Web-based sources, Google had the best combination of sensitivity and PPV in detecting Bayesian change points in influenza-related data streams. Findings demonstrated that change points in Google Flu Trends, Twitter, and Wikipedia data occasionally aligned well with change points captured in CDC ILI data, yet these sources did not detect all changes in CDC data and should be further studied and developed.

Tran A, Trevennec C, Lutwama J, Sserugga J, Gély M, Pittiglio C, Pinto J, Chevalier V

Development and Assessment of a Geographic Knowledge-Based Model for Mapping Suitable Areas for Rift Valley Fever Transmission in Eastern Africa

PLoS Negl Trop Dis 2016;10(9)

Rift Valley fever (RVF), a mosquito-borne disease affecting ruminants and humans, is one of the most important viral zoonoses in Africa. The RVF virus (RVFV) is transmitted from ruminant to ruminant by mosquitoes. Different climatic, environmental, and socio-economic factors may impact the transmission of the virus. The objective of the present study was to develop a geographic knowledge-based method to map the areas suitable for RVF amplification and RVF spread in four East African countries, namely, Kenya, Tanzania, Uganda, (three countries which have been historically affected by RVF), and Ethiopia (where the disease has never been reported but which shares borders with infected countries), and to assess the predictive accuracy of the model using livestock outbreak data from Kenya and Tanzania. Risk factors and their relative importance regarding RVF amplification and spread were identified from a literature review. The data were imported into a geographic information system (GIS) and processed to produce standardized spatial risk factor layers, namely a mosquito index (suitability for RVF mosquito vectors), sheep density, goat density, cattle density, proximity to markets, road density, railways density, proximity to water bodies, proximity to wildlife national parks. A numerical weight was calculated for each risk factor using an analytical hierarchy process. The corresponding geographic data were collected, standardized, and combined based on a weighted linear combination to produce maps of the suitability for RVF transmission. The accuracy of the resulting maps was assessed using RVF outbreak locations in livestock reported in Kenya and Tanzania between 1998 and 2012 and the ROC curve analysis. Results confirmed the capacity of the geographic information system-based multi-criteria evaluation method to synthesize available scientific knowledge and to accurately map (AUC = 0.786; 95% CI [0.730–0.842]) the spatial heterogeneity of RVF suitability in East Africa. Some areas may be at-risk without having experienced outbreaks in past years. The identification of these areas is essential for implementing risk-based surveillance and reducing the impact of RVF human and animal outbreaks in the coming years (until 2016, Uganda and Ethiopia remained free from outbreaks, but these two countries are highly vulnerable to the disease). This approach provides users with a straightforward update of the maps according to data availability and contributes to the further development of scientific knowledge.