Modelling, Simulation and Social Network Data: What's New for Public Health and Epidemiology Informatics?

L. Toubiana[1], N Griffon[1,2], Section Editors for the IMIA Yearbook Section on Public Health and Epidemiology Informatics
[1] INSERM UMRS 1142 LIMICS, Université Pierre et Marie Curie, Paris, France
[2] CISMeF, TIBS, LITIS EA 4108, Rouen University Hospital, Normandy, Rouen, France

Summary

Objectives: Summarize excellent current research in the field of Public Health and Epidemiology Informatics.

Method: Synopsis of the articles selected for the IMIA Yearbook 2015.

Results: Four papers from international peer-reviewed journals have been selected as best papers for the section on Public Health and Epidemiology Informatics.

Conclusions: The selected articles illustrate current research regarding the impact and assessment of health IT and the latest developments in health information exchange.

Keywords

Public health, epidemiology, medical informatics, International Medical Informatics Association, health information systems, hospital information systems

Introduction

One objective of the IMIA Yearbook is to review research articles published the previous year in peer-reviewed journals to identify research trends and excellent research papers in the different fields of medical informatics, reflecting a broad spectrum of topics.

Until this year, papers on public health and epidemiology informatics were included in the section Health Information Systems (HIS). Several high quality papers on public health and epidemiology informatics used to be selected as candidate best papers of the HIS section. However, as they were less central to the topic of the section, it was difficult to keep them as best papers. As a consequence, papers related to public health were poorly represented, in spite of their quality, of their growing number, and of the diversity of the issues treated. This was particularly unfortunate since the field of public health and epidemiology informatics has a tremendous potential of attractiveness, and it starts to be the theme of emergence and fertilisation for many important methodological aspects of medical informatics.

Public health Informatics was introduced in 2003 in the MeSH thesaurus, and defined as "the systematic application of information and computer sciences to public health practice, research, and learning". Epidemiology Informatics concerns the use of computer science and the considerable amount of data that computer science has made available in order to enhance epidemiological knowledge. The combination of Public Health Informatics and Epidemiology informatics has led to a new section of the IMIA Yearbook in 2015. This new section addresses any issues related to the use of informatics or computer science in the field of epidemiology and public health, except the use of computational power to run complex models.

During the review of 2014 publications, we observed that authors in this field tried to gather an increasing number of data to improve the sensitivity of surveillance systems. Two trends were identified regarding the aim to ensure interoperability between already existing data sources, and the use of the huge amount of data created/collected on the web.

Research work also focused on mathematic modeling of health processes. For example, the impact of climate change and the increasing weight of social media need a better understanding in order to anticipate.

Finally, we found that public health information systems, whether related to entire populations or to specific events (e.g. outbreaks), are of interest to many researchers. The pre-selected publications mostly concern this subfield.

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 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 were excluded from our search. The study was performed at the beginning of January 2015, and the search over the year 2014 returned a total of 1,371 references. The query is available online.

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 62 references that were retained by at least one reviewer or classified as "pending" by both of them.

The two section editors jointly reviewed the 62 references and drafted a consensual list of 16 candidate best papers. All pre-selected 16 papers were then peer-reviewed by editors and external reviewers (at least four reviewers per paper). Four 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 [1] describe the whole selection process.

Table 1: Best papers selected for the IMIA Yearbook of Medical Informatics 2015 in the section 'Public Health and Epidemiology Informatics'. The articles are listed in alphabetical order of the first author's surname.

Conclusion and Outlook

The selection of best papers from the large corpora of published studies was performed according to criteria of quality, regardless of the 2014 theme of the Yearbook ("Patient-centered care coordination").

The paper by Dixon et al. [2] advocates for an increased interoperability between data sources. Interoperability is one solution for public health agencies to monitor National health status. Mixing multiple data sources may allow the detection of signals that would be too weak to detect with fewer data. The paper describes the difficulties encountered due to the lack of semantic interoperability when setting up surveillance. Authors discuss potential strategies to overcome this issue and finally

Table 1 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2014 in the section 'Public Health and Epidemiology Informatics'. The articles are listed in alphabetical order of the first author's surname.

Section
Public Health and Epidemiology Informatics
■ Barboza P, Vaillant L, Le Strat Y, Hartley DM, Nelson NP, Mawudeku A, Madoff LC, Linge JP, Collier N, Brownstein JS, Astagneau P: Factors influencing performance of internet-based biosurveillance systems used in epidemic intelligence for early detection of infectious diseases outbreaks. PLoS One 2014;9:e90536.
■ Khormi HM, Kumar L: Climate change and the potential global distribution of Aedes aegypti: spatial modelling using GIS and CLIMEX. Geospat Health 2014;8:405–15.
■ Pawelek KA, Oeldorf-Hirsch A, Rong L: Modeling the impact of twitter on influenza epidemics. Math Biosci Eng 2014;11:1337–56.
■ Rutter P, Mytton O, Ellis B, Donaldson L: Access to the NHS by telephone and Internet during an influenza pandemic: an observational study. BMJ Open 2014;4:e004174.

plead for a cooperative approach between data senders and public health agencies to converge toward standards.

A second solution to monitor National health status is to use the huge amount of data generated by Internet users. The idea of using web-generated data emerges in the early 2000s, allowing earlier signal detection. This year again, several articles addressed research issues in this field with different data source. Generous et al. [3] used Wikipedia's log data, Aslam et al. [4] used twitter data, and Yom-Tov et al. [5] used both twitter and Microsoft Bing search engine data. The approaches used were slightly different. Generous et al. [3] used Wikipedia's log data to detect epidemics, assuming that the beginning of an epidemic may cause an increase in the view of the disease page. The main difficulty encountered is the localization of the user. To overcome this, the authors make an audacious hypothesis based on the intersection of language (that may be specific of a particular country) and disease (that may be observed in very few countries). Aslam et al. [4] investigated the usefulness of Twitter for flu surveillance. They used an automated classifier to determine the relevance of a tweet. The originality of the work by Yom-Tov et al. [5] is to focus on the detection of epidemics following mass-gathering events. Such events have indeed the potential for a wide dissemination and are therefore important to monitor. It should be noticed that all these publications lacked a convincing assessment. The absence of a gold standard can be an acceptable reason,

but even if such a gold standard exists, the authors' methods to evaluate their tools were weak: crude correlation coefficient, use of all the data for training, without a test set for evaluation.

Barboza et al. [6] did not develop a new system but they evaluated the performances of already existing biosurveillance systems. Signals detected were compared to a gold standard based on investigations by the French institute for public health surveillance (InVS). Even if the gold standard may be questionable (as is any expertise-based gold standard), and the specificity was not evaluated, this study allows a proper evaluation of biosurveillance systems and an exploratory analysis about what works and what does not, which explains why this paper was finally selected as a best paper.

Five papers, Khormi et al. [8], Nikakhtar et al. [9], Pawelek et al. [7], Gupta et al. [10], Jandarov et al. [11] used new approaches to model epidemics or their evolution by using different parameters. These authors used informatics either by introducing it in the model or by using it to run the model. The first selected paper, by Pawelek [7], offers an interesting approach to evaluate the impact of web 2.0 on the dynamics of epidemics. The second, by Khormi [8], combines geography, epidemics and climate changes. Nikakhtar et al. [9] describe an approach to model costs and effects of the management of malaria due to Plasmodium falciparum in sub-Saharan Africa. Gutpa et al. [10] describe a scalable computational framework for establishing

long-term behaviour of stochastic reaction networks. Jandarov et al [11] use a gravity model to infer the spatiotemporal dynamics of infectious diseases.

Rutter et al. [12] proposed a detailed description of a service provided to limit the impact of influenza pandemic on the British Health System. In short, a telephone and Internet-based national service was set up. Any person with flu-like symptoms could access it. Users were either advised to call an ambulance, to visit a general practitioner or a center for antiviral treatment. They could also be simply reassured. There are some safety issues, and one could have expected the service to be evaluated. Nevertheless, the novelty and the scale of the service made us selecting it as a best paper. Such a service may help avoid inundating primary care and prevent disease transmission, while it contributes to monitor the disease in the general population.

Verdot et al. [13] worked on the health status surveillance of prisoners, a particularly understudied and under-monitored population. This literature review describes surveillance systems in 15 countries. With two exceptions (countries with standardized and routinely computerized health data), it appears that data are sparse in most countries: crude mortality data, episodic epidemiologic study…. There is, at best, a surveillance of infectious diseases.

Acknowledgements

Correspondence to:
Dr. Nicolas Griffon
Unité d'Informatique Médicale
CHU de Rouen
1 rue de Germont
76031, Rouen
Tel. +33 6 42 25 44 11
E-mail: nicolas.griffon@chu-rouen.fr

Dr. Laurent Toubiana, PhD
INSERM UMRS 1142 "LIMICS"
15, rue de l'École de Médecine
75006 Paris, France
Tel: +33 1 44 27 91 97
E-mail: Laurent.toubiana@inserm.fr

Appendix: Content Summaries of Selected Best Papers for the IMIA Yearbook 2015, Section 'Public Health and Epidemiology Informatics'

Barboza P, Vaillant L, Le Strat Y, Hartley DM, Nelson NP, Mawudeku A, Madoff LC, Linge JP, Collier N, Brownstein JS, Astagneau P

Factors influencing performance of internet-based biosurveillance systems used in epidemic intelligence for early detection of infectious diseases outbreaks

PLoS One 2014;9:e90536

The capacity of biosurveillance systems to detect an entire epidemic event has not yet been fully elucidated. In this paper, authors propose a gold standard allowing for such an evaluation, and use it to compare 6 biosurveillance systems that have already been published. The gold standard is extracted from the "French institute for public health surveillance" (InVS) publication, the weekly international epidemiological bulletin (BHI). Such a gold standard is not exhaustive, but allows for the analysis of sensitivity. The outputs of biosurveillance systems were searched for signals corresponding to gold standard events and multiple metrics were computed. Authors also tried to identify determinants of good sensitivity. The gold standard included 86 events for which the tools emitted 472 signals. The capacity of the tools to detect events was high (ranging from 83% to 95%). However, when considering only the detection prior to the gold standard, there is a ten-point decrease in sensitivity (range from 71% to 85%). It appears that moderated systems perform better than automated systems, and the most important source of information is the media, outranking official reports. The language is also an important factor to take into account. The main result of this evaluation is probably the importance of synergies between tools. Nevertheless, the false positive signals were not evaluated. It might counterbalance the benefits of synergies. There is a huge need for an internationally validated gold standard allowing further evaluation.

Khormi HM, Kumar L

Climate change and the potential global distribution of Aedes aegypti: spatial modelling using GIS and CLIMEX

Geospat Health 2014;8:405–15

Authors studied the distribution area of a Dengue vector mosquito according to climate change scenarios, and the resulting risk for public health. They used a model allowing for simulations according to different climate change scenarios over time spans of 15 and 55 years. Simulations showed that many areas could become favourable to the mosquito species concerned. This paper describes a comprehensive modelling approach for the impact of climate change. It is the only paper of our selection that addresses this issue that is nonetheless central to the discussions in epidemiology today. The paper is very well documented and is supported by numerous data.

Pawelek KA, Oeldorf-Hirsch A, Rong L

Modeling the impact of twitter on influenza epidemics

Math Biosci Eng 2014;11:1337–56

Pawelek et al. propose an interesting approach to evaluate the impact of web 2.0 on epidemic dynamics. The authors describe a model for the influenza epidemic, not by assessing the epidemic directly, but by assessing its effect on behaviour. The paper answers the original question regarding the impact of data circulating on social networking sites (Twitter) on the dynamics of the epidemic. Based on a classic compartment model, this is a very high-quality and well-documented theoretical study. Sensitivity analysis was performed by comparing the model to the real data recorded in England during the H1N1 outbreak. The result is a novel example of using data collected on social networks for the detection and monitoring of seasonal epidemics.

Rutter P, Mytton O, Ellis B, Donaldson L
Access to the NHS by telephone and Internet during an influenza pandemic: an observational study
BMJ Open 2014;4:e004174

To avoid long waiting lines in front of medical centres – an excellent location for disease transmission - a national telephone and Internet-based service was established, coupled to a network of antiviral drug distribution centres during the 2009-2010 influenza pandemic. This paper describes the service and reports on its use by the British population. All persons exhibiting 'flu-like' symptoms could access the service for more information. An algorithm automatically redirected patients needing urgent medical attention to a physician. The next steps determined whether the patient should be authorised to obtain antiviral medication in a distribution centre, advised to contact other healthcare providers, or simply reassured. During the 6-month activation period, 2.73 million unique patients used the service. There was a clear superposition between the use rate of the service and GP consultations for influenza-like illness, with a 1 to 6 ratio for the service. Some 9% of patients were redirected and advised to seek emergency care. Of the 2,488,504 remaining patients, 38% were advised to visit their GP, and 72%

were authorised to obtain antiviral medication - or both. Only 64% of the latter group went to a distribution centre to get their medication. Internet and telephone accesses were almost equally used. This was probably one of the first national systems in the world that involved only non-medically trained people as the primary interface with patients, and provided a drug that normally requires medical prescription. According to GPs, this system probably reduced the pressure on GPs at a very busy time. Some problems remain, including the probable lack of specificity, which may increase the emergence of resistance to the medication and the possible lack of sensitivity that may lead to dramatic events. Nevertheless, such an initiative may be of interest to other countries for pandemic contingency planning.

References

1. Lamy JB, Séroussi B, Griffon N, Kerdelhué G, Jaulent MC, Bouaud J. Toward a formalization of the process to select IMIA Yearbook best papers. Methods Inf Med 2015;54(2):135-44.
2. Dixon BE, Vreeman DJ, Grannis SJ. The long road to semantic interoperability in support of public health: experiences from two states. J Biomed Inform 2014;49:3–8.
3. Generous N, Fairchild G, Deshpande A, Del Valle SY, Priedhorsky R. Global disease monitoring and forecasting with Wikipedia. PLoS Comput Biol 2014;10:e1003892.
4. Aslam AA, Tsou M-H, Spitzberg BH, An L, Gawron JM, Gupta DK, et al. The reliability of tweets as a supplementary method of seasonal influenza surveillance. J Med Internet Res 2014;16:e250.
5. Yom-Tov E, Borsa D, Cox IJ, McKendry RA. Detecting disease outbreaks in mass gatherings using Internet data. J Med Internet Res 2014;16:e154.
6. Barboza P, Vaillant L, Le Strat Y, Hartley DM, Nelson NP, Mawudeku A, et al. Factors influencing performance of internet-based biosurveillance systems used in epidemic intelligence for early detection of infectious diseases outbreaks. PLoS One 2014;9:e90536.
7. Pawelek KA, Oeldorf-Hirsch A, Rong L. Modeling the impact of twitter on influenza epidemics. Math Biosci Eng 2014;11:1337–56.
8. Khormi HM, Kumar L. Climate change and the potential global distribution of Aedes aegypti: spatial modelling using GIS and CLIMEX. Geospat Health 2014;8:405–15.
9. Nikakhtar A, Hsiang SM. Incorporating the dynamics of epidemics in simulation models of healthcare systems. Simul Model Pract Theory 2014;43:67–78.
10. Gupta A, Briat C, Khammash M. A scalable computational framework for establishing long-term behavior of stochastic reaction networks. PLoS Comput Biol 2014;10:e1003669.
11. Jandarov R, Haran M, Bjørnstad O, Grenfell B. Emulating a gravity model to infer the spatiotemporal dynamics of an infectious disease. J R Stat Soc Ser C (Applied Stat 2014;63:423–44.
12. Rutter P, Mytton O, Ellis B, Donaldson L. Access to the NHS by telephone and Internet during an influenza pandemic: an observational study. BMJ Open 2014;4:e004174.
13. Verdot C, Godin-Blandeau E, Grémy I, Develay A-E. Monitoring systems and national surveys on prison health in France and abroad. Eur J Public Health 2015;25:167–72.