

# Public Health and Epidemiology Informatics: Recent Research Trends

Gayo Diallo<sup>1</sup>, Georgeta Bordea<sup>1,2</sup>, Section Editors for the IMIA Yearbook Section on Public Health and Epidemiology Informatics

<sup>1</sup> INRIA SISTM, Team ERIAS - INSERM Bordeaux Population Health Research Center, Univ. Bordeaux, Bordeaux, France

<sup>2</sup> Team ERIAS - INSERM BPH Research Center & LaBRI UMR 5800, Univ. Bordeaux, Bordeaux, France

## Summary

**Objectives:** To introduce and analyse current trends in Public Health and Epidemiology Informatics.

**Methods:** PubMed search of 2020 literature on public health and epidemiology informatics was conducted and all retrieved references were reviewed by the two section editors. Then, 15 candidate best papers were selected among the 920 references. These papers were then peer-reviewed by the two section editors, two chief editors, and external reviewers, including at least two senior faculty, to allow the Editorial Committee of the 2021 International Medical Informatics Association (IMIA) Yearbook to make an informed decision regarding the selection of the best papers.

**Results:** Among the 920 references retrieved from PubMed, four were suggested as best papers and the first three were finally selected. The fourth paper was excluded because of reproducibility issues. The first best paper is a very public health focused paper with health informatics and biostatistics methods applied to stratify patients within a cohort in order to identify those at risk of suicide; the second paper describes the use of a randomized design to test the likely impact of fear-based messages, with and without empowering self-management elements, on patient consultations or antibiotic requests for influenza-like illnesses. The third selected paper evaluates the perception among communities of routine use of Whole Genome Sequencing and Big Data technologies to capture more detailed and specific personal information.

**Conclusions:** The findings from the three studies suggest that using Public Health and Epidemiology Informatics methods could leverage, when combined with Deep Learning, early interventions and appropriate treatments to mitigate suicide risk. Further, they also demonstrate that well informing and empowering patients could help them to be involved more in their care process.

## Keywords

Public health, epidemiology informatics, IMIA Yearbook

Yearb Med Inform 2021:280-2

<http://dx.doi.org/10.1055/s-0041-1726530>

## 1 Introduction

The trend observed in recent years regarding the increased use of digital technology in healthcare and the recent evolution of artificial intelligence (AI) and deep learning technologies continue to have a positive impact on population-based studies.

A review of research published last year showed that the increasing availability of data from either the internet and social networks or more traditional health data sources such as Electronic Health Records (EHRs) is fueling research on topics covering statistics, health informatics, and epidemiology [1]. These topics continue to be key components of Public Health Data Science.

The current synopsis reviews the scientific literature published in 2020 in the domain of medical informatics with a focus on the sub-field of public health and epidemiology. The synopsis aims to identify new topics and trends within public health as compared to previous years. Our review further describes the selection process followed to choose the best papers that were published in 2020. To this end, the section editors coordinated the reviewing process of the most representative research works, based on quality and originality.

## 2 Methods

A comprehensive literature search was performed using PubMed/MEDLINE database from the National Center for Biotechnology Information (NCBI). A large set of Medical

Subject Headings (MeSH) descriptors was used to retrieve relevant studies. The queries targeted public health or epidemiological journal articles over the year 2020 that included medical informatics topics.

Returned references addressing topics of the other sections of the Yearbook, e.g., those related specifically to pandemics and Corona Virus (COVID), were excluded from our search. Further, a cross-check is performed with the 2021 International Medical Informatics Association (IMIA) Yearbook Special Topic “Managing Pandemics with Health Informatics: successes and challenges” section for any overlaps in themes that are not obvious.

The study was performed at the beginning of February 2021, and the search returned a total of 920 references. Articles were separately reviewed by the two section editors considering in the process the number of citations of each reference. Articles were first classified into three categories: keep, discard, or pending using the BibReview tool [2]. Then, the two lists of references were merged, yielding 96 references that were retained by at least one reviewer or classified as “pending” by both of them. The two section editors jointly reviewed 96 references Special Topic section was performed in order to double-check if any overlap between the two sections exists.

All of these papers were then peer-reviewed by editors and external reviewers. Each paper was reviewed by at least five reviewers. Four papers were suggested at the end of this process. However, the fourth paper [3], although interesting for the field, was removed due to reproducibility issue.

Indeed, the described study use a proprietary tool and the dataset is not available for reuse. Finally, three papers, respectively, [4], [5] and [6] were selected as best papers by the Yearbook Editorial Committee (Table 1). A content summary of these selected papers can be found in the appendix of this synopsis.

### 3 Results

Beyond the high-quality of research publications and their contributions, several research themes were prioritized based on their potential impact for future research. These themes cover diverse aspects including the interpretation of deep-learning approaches, assessing benefits and risks of using linked data in disease surveillance, open access aspects of epidemiology datasets, addressing gender bias in medical informatics, and optimisation of messages in online public health campaigns. These are still open but pressing research questions that bear witness to sometimes conflicting goals in public health informatics that are related to access to Big Data, trust in artificial intelligence, privacy, minority bias, and a general shift to online investigation.

Overall, the findings from the three selected studies suggest that using Public Health and Epidemiology Informatics methods could leverage, when combined with Deep Learning, early interventions and appropriate treatments to mitigate suicide risk. Further, they also demonstrate that when people are well-informed, here using a jury proxy [6], they are likely to support routine collection, linkage and use of administrative and pathogenomic data for the purposes of public health research. In addition, empowering patients could help them to be involved more in their care process.

**Table 1** Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2021 in the section 'Public Health and Epidemiology Informatics'. The articles are listed in alphabetical order of the first author's surname.

Section
<b>Public Health and Epidemiology Informatics</b>
<ul style="list-style-type: none"> <li>▪ Degeling C, Carter SM, van Oijen AM, McAnulty J, Sintchenko V, Braunack-Mayer A, Yarwood T, Johnson J, Gilbert GL. Community perspectives on the benefits and risks of technologically enhanced communicable disease surveillance systems: a report on four community juries. <i>BMC Med Ethics</i> 2020;21(1):31.</li> <li>▪ Roope LSJ, Tonkin-Crine S, Herd N, Michie S, Pouwels KB, Castro-Sanchez E, Sallis A, Hopkins S, Robotham JV, Crook DW, Peto T, Peters M, Butler CC, Walker AS, Wordsworth S. Reducing expectations for antibiotics in primary care: a randomised experiment to test the response to fear-based messages about antimicrobial resistance. <i>BMC Med</i> 2020;18(1):110.</li> <li>▪ Zheng L, Wang O, Hao S, Ye C, Liu M, Xia M, Sabo AL, Markovic L, Stearns F, Kanov L, Sylvester KL, Widen R, McElhinney DB, Zhang W, Liao J, Ling XB. Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. <i>Transl Psychiatry</i> 2020;10(1):72.</li> </ul>

### 4 Conclusion

Our analysis confirmed once again the general trend of increased interest in Public Health and Epidemiology Informatics (PHEI) in recent years, even beyond the specific context of an ongoing pandemic. Access to big data, trust in artificial intelligence, privacy, and minority bias are some of the most pressing research questions in this area with a shift to online investigation that empowers patients to be more involved in their care process.

#### Acknowledgements

The PHEI section editors would like to thank the reviewers for their valuable contribution in evaluating and selecting the best papers for this section of the 2021 IMIA Yearbook.

#### References

1. Cossin S, Rodolphe Thiébaud R. Public Health and Epidemiology Informatics: Recent Research Trends Moving toward Public Health Data Science. *Yearb Med Inform* 2020; 29(1):231–4.
2. Lamy JB, 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):135–44.

3. Eletreby R, Zhuang Y, Carley KM, Yagan O, Poor HV. The effects of evolutionary adaptations on spreading processes in complex networks. *Proc Natl Acad Sci U S A* 2020;117(11):5664–70.
4. Zheng L, Wang O, Hao S, Ye C, Liu M, Xia M, et al. Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. *Transl Psychiatry* 2020;10(1):72.
5. Roope L, Tonkin-Crine S, Herd N, Michie S, Pouwels K, Castro-Sánchez E, et al. Reducing expectations for antibiotics in primary care: a randomised experiment to test the response to fear-based messages about antimicrobial resistance. *BMC Med* 2020;18(1):110.
6. Degeling C, Carter S, Oijen A, Mc Anulty J, Sintchenko V, Braunack-Mayer A, et al. Community perspectives on the benefits and risks of technologically enhanced communicable disease surveillance systems: A report on four community juries. *BMC Med Ethics* 2020;21(1):31.

#### Correspondence to:

Gayo Diallo  
INRIA SISTM, Team ERIAS  
INSERM Bordeaux  
Population Health Research Center  
Univ. Bordeaux  
146, rue Léo Saignat  
F-33000 Bordeaux, France  
E-mail: gayo.diallo@u-bordeaux.fr

## Appendix: Content Summary of Selected Best Papers 2021 IMIA Yearbook Section Public Health and Epidemiology Informatics

Zheng L, Wang O, Hao S, Ye C, Liu M, Xia M, Sabo AL, Markovic L, Stearns F, Kanov L, Sylvester KL, Widen R, McElhinney DB, Zhang W, Liao J, Ling XB

**Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records**

*Transl Psychiatry* 2020;10(1):72

The paper of Zheng et al. reports a study focusing on tackling suicide attempt. It is a public health focused paper combining health informatics and biostatistics methods and which are applied to stratify patients within a cohort in order to identify those at risk of suicide. They developed an early-warning system for high-risk suicide attempt patients through the design and implementation of a population-based risk stratification surveillance system. A Deep Neural Network (DNN) model was trained for the prediction, stratification, and calibration. Then a Local Interpretable Model-agnostic Explanations algorithm was utilised to interpret the risk stratification results. As of result, a total of 117 features were significant in the predictive model; the DNN on the EHR-based data, enabled finding that suicide attempts patients were more likely to be gin: age groups of 6–54,

diagnosed mental health conditions or pain, to have suicide attempts, treated by psychotropic medications, and have open wounds or injuries due to unspecific reasons. The findings of the study enable early interventions and appropriate treatments to mitigate suicide risk.

Roope LSJ, Tonkin-Crine S, Herd N, Michie S, Pouwels KB, Castro-Sanchez E, Sallis A, Hopkins S, Robotham JV, Crook DW, Peto T, Peters M, Butler CC, Walker AS, Wordsworth S

**Reducing expectations for antibiotics in primary care: a randomised experiment to test the response to fear-based messages about antimicrobial resistance**

*BMC Med* 2020;18(1):110

The aim of this study is to test the likely impact of fear-based messages, with and without empowering self-management elements, on patient consultations or antibiotic requests for influenza-like illnesses, using a randomised design. To do so, they relied on the use of an on-line survey of adult members of the UK general public. They proceed to the randomisation of patients, to receive one of three different messages about antibiotics and Antimicrobial Resistance. As of results, they have been able to demonstrate that empowering patient could help to better involve them in their care process and make a better decision regarding Antibiotics related issues. The study concludes that while fear-only messages could be effective in public campaigns to reduce inappropriate antibiotic use, they should be combined with messages empowering patients to self-manage symptoms effectively without antibiotics.

Degeling C, Carter SM, van Oijen AM, McNulty J, Sintchenko V, Braunack-Mayer A, Yarwood T, Johnson J, Gilbert GL

**Community perspectives on the benefits and risks of technologically enhanced communicable disease surveillance systems: a report on four community juries**

*BMC Med Ethics* 2020;21(1):31

The paper focuses on assessing how routinely using Whole Genome Sequencing (WGS) and Big Data technologies to capture more detailed and specific personal information could be perceived among communities in two demographically different Sydney municipalities and two regional cities in New South Wales, Australia (western Sydney, Wollongong, Tamworth, eastern Sydney). Four community juries were created by recruiting participants at each study site by an independent professional research service. The aim was to elicit the views of well-informed community members on the acceptability and legitimacy of making pathogen WGS and linked administrative data available for public health research using this information in concert with data linkage and machine learning to enhance communicable disease surveillance systems. Participants across all four events strongly supported the introduction of data linkage and pathogenomics to public health research under current research governance structures. This study demonstrates that when public is well informed, here a jury proxy is used, they are likely to support routine collection, linkage and use of administrative and pathogenomics data for the purposes of public health research.