JMIR Medical Informatics

Impact Factor (2023): 3.1
Volume 8 (2020), Issue 2 ISSN 2291-9694 Editor in Chief: Christian Lovis, MD, MPH, FACMI

Contents

Original Papers

Original Paper

Evaluating the Applications of Health Information Technologies in China During the Past 11 Years: Consecutive Survey Data Analysis

Jun Liang[1], MS; Ying Li[2], MD; Zhongan Zhang[3], BS; Dongxia Shen[4], MS; Jie Xu[1], MS; Gang Yu[5], MS; Siqi Dai[6], MSc; Fangmin Ge[7], MS; Jianbo Lei[8,9,10], MD, PhD

[1]IT Center, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China

[2]Department of Burn and Plastic Surgery, Affiliated Hospital of Southwest Medical University, Luzhou, China

[3]Performance Management Department, Qingdao Central Hospital, Qingdao, China

[4]Editorial Department of Journal of Practical Oncology, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China

[5]IT Center, Children's Hospital, School of Medicine, Zhejiang University, Hangzhou, China

[6]School of Medicine, Zhejiang University, Hangzhou, China

[7]International Network Medical Center, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China

[8]Center for Medical Informatics, Peking University, Beijing, China

[9]Institute of Medical Technology, Health Science Center, Peking University, Beijing, China

[10]School of Medical Informatics and Engineering, Southwest Medical University, Luzhou, China

Corresponding Author:
Jianbo Lei, MD, PhD
Institute of Medical Technology
Health Science Center
Peking University
38 Xueyuan Rd, Haidian District
Beijing, 100191
China
Phone: 86 8280 5901
Email: jblei@hsc.pku.edu.cn

Abstract

Background: To achieve universal access to medical resources, China introduced its second health care reform in 2010, with health information technologies (HIT) as an important technical support point.

Objective: This study is the first attempt to explore the unique contributions and characteristics of HIT development in Chinese hospitals from the three major aspects of hospital HIT—human resources, funding, and materials—in an all-around, multi-angled, and time-longitudinal manner, so as to serve as a reference for decision makers in China and the rest of the world when formulating HIT development strategies.

Methods: A longitudinal research method is used to analyze the results of the CHIMA Annual Survey of Hospital Information System in China carried out by a Chinese national industrial association, CHIMA, from 2007 to 2018. The development characteristics of human resources, funding, and materials of HIT in China for the past 12 years are summarized. The Bass model is used to fit and predict the popularization trend of EMR in Chinese hospitals from 2007 to 2020.

Results: From 2007 to 2018, the CHIMA Annual Survey interviewed 10,954 hospital CIOs across 32 administrative regions in Mainland China. Compared with 2007, as of 2018, in terms of human resources, the average full time equivalent (FTE) count in each hospital's IT center is still lower than the average level of US counterparts in 2014 (9.66 FTEs vs. 34 FTEs). The proportion of CIOs with a master's degree or above was 25.61%, showing an increase of 18.51%, among which those with computer-related backgrounds accounted for 64.75%, however, those with a medical informatics background only accounted for 3.67%. In terms of funding, the sampled hospitals' annual HIT investment increased from ¥957,700 (US $136,874) to ¥6.376 million (US $911,261), and the average investment per bed increased from ¥4,600 (US $658) to ¥8,100 (US $1158). In terms of information system construction, as of 2018, the average EMR implementation rate of the sampled hospitals exceeded the average level of their US

counterparts in 2015 and their German counterparts in 2017 (85.26% vs. 83.8% vs. 68.4%, respectively). The results of the Bass prediction model show that Chinese hospitals will likely reach an adoption rate of 91.4% by 2020 (R^2=0.95).

Conclusions: In more than 10 years, based on this top-down approach, China's medical care industry has accepted government instructions and implemented the unified model planned by administrative intervention. With only about one-fifth of the required funding, and about one-fourth of the required human resources per hospital as compared to the US HITECH project, China's EMR coverage in 2018 exceeded the average level of its US counterparts in 2015 and German counterparts in 2017. This experience deserves further study and analysis by other countries.

(JMIR Med Inform 2020;8(2):e17006) doi:10.2196/17006

KEYWORDS

medical informatics; health information technologies; China; health care reform; hospitals

Introduction

Health information technologies (HIT) can effectively improve the quality and efficiency of medical services, distribution of health care resources, safety in health care, and output of scientific research. Therefore, governments of various countries have set up ambitious plans to develop HIT and invested enormous amounts of money in this development, using HIT as an important starting point for the reformation of medical services and medical systems.

The US government invested $787 billion in the American Recovery and Reinvestment Act of 2009. In particular, $19 billion of this investment was used to promote nationalized and interoperable health information systems and implement them through the Health Information Technology for Economic and Clinical Health (HITECH) Act [1]. Its core Meaningful Use strategy has achieved initial results [2]. As of 2014, approximately 75.5% of US hospitals had at least a basic system with a defined set of functions applied in at least one hospital unit. About 69% of these hospitals supported the exchange of laboratory examination results, 65% supported exchange of radiological examination reports, 64% supported exchange of progress notes, and 55% supported exchange of medication histories, compared with 35%, 37%, 25%, and 21%, respectively, in 2008 [3]. The United Kingdom launched the National Programme for IT in 2005. By 2011, the utilization rate of electronic health records (EHR) for primary care was close to 100% [4], and the successful experience of the US HITECH Act was further introduced in 2014 [4].

China has been no exception to this trend. As early as the beginning of the second health care reform in 2010, the government adopted HIT as one of the "four beams and eight pillars" supporting health care reform [5] and successively promulgated 31 national policies and 134 technical standards covering all aspects of hospital, population health, and medical security system digitalization.

In order to build the HIT system, as detailed in the Healthy China 2020: Strategic Research Report released by the National Health Commission of the People's Republic of China in 2012 [6], a national budget of US $10 billion will be invested to build the National Electronic Health Information System Project by 2020, more than one-seventh of the total investment of US $68 billion designated for the plan. As of 2015, the central government had actually invested more than US $3.5 billion. For details of the investment and expected results, see Multimedia Appendix 1. According to the latest administrative directive issued by the National Health Commission of the People's Republic of China in August 2018, the use of electronic medical records (EMR) in hospitals should be included in the index system for hospital performance evaluation [7].

Despite the formulation of very active macro policies and the investment of a large amount of funds, governments of various countries have always faced significant challenges in the technological research and development, project implementation, effect evaluation, and speed of advancement of HIT. Governments, academic circles, and industries have constantly presented the relevant experience and lessons. Kruse et al [8] collected 3636 articles and selected 37 articles for final research; they found that 81% of the research projects believed that the HIT projects already implemented had a positive effect on the quality and cost of medical care. Gold et al [3] advanced the claim that although HITECH provides administrative and economic resources for the standards and interoperability of EHRs and HIT, the law does not stipulate how to achieve them. The US administrative system retains considerable autonomy for the private sector, making it even more difficult to reach a consensus under the current situation of relatively independent public power at the federal and state levels. This has led to a substantial delay in the implementation of HITECH. At present, it is too early to evaluate the final effect of HIT projects implemented between 2009 and 2015. Adler-Milstein et al [9] found that with the stimulation of HITECH, as of 2013, EHRs have been used in more than 50% of hospitals, with some regional differences; rural and small specialized hospitals lag far behind, potentially leading to problems of medical resource allocation.

As the largest country in the world in terms of population and number of hospitals and the second largest in total economic volume, China currently lacks relevant research on the application status, characteristics, and challenges of HIT in its hospitals. In this study, we try to answer the following questions:

- How can we describe, evaluate, and summarize the achievements and problems in China's HIT development from 2007 to 2018?
- During this period, compared with countries with advanced HIT such as the United States, what are China's characteristics in terms of the number and quality of HIT employees, capital and resource investment, network

support environment, and application of clinical information systems (CIS) such as EMR?

Methods

Data Resources

Our data are from the 2007-2018 China Hospital Information Management Association (CHIMA) annual survey hospital information systems, which is the only national HIT industry survey covering a period of more than 10 years in China. Over the last decade or so, CHIMA [10] has used the questionnaire issued by the journal Chinese Digital Medicine to conduct continuous research on China's HIT application market in March of every year. The research area covered 34 administrative regions of mainland China. The institutions reviewed included general hospitals, specialized hospitals, traditional Chinese medicine hospitals, and integrated traditional Chinese and western medicine hospitals. The interviewees were chief information officers (CIOs) who were responsible for the information technology (IT) departments of the hospitals. The research method was designed with reference to the Healthcare Information and Management Systems Society (HIMSS) annual survey in the United States, and hospitals that did not respond in time received email and telephone notifications.

The CHIMA survey comprised 9 parts: respondents' basic information, IT application, infrastructure and hardware use, information system application, IT outsourcing, IT construction obstacles, information system construction investment, data standardization, and regional medical and health information system construction. We mainly used the data from the first to seventh of the 9 parts; in particular, the data from parts I-V and VII: respondents' basic information, IT application, infrastructure and hardware use, information system application, IT outsourcing, and information system construction investment. Each year’s survey report provides a summary of the current situation of hospital digitalization and the overall trend of HIT in China. The 2018 survey was completed between March 2019 and June 2019 and released on September 10, 2019.

Research Subjects: Hospital Information Technology Department–Related Attributes of the China Hospital Information Management Association Annual Survey

In China, most hospitals purchase HIT software from the HIT market, which is outsourced by system suppliers. Therefore, the IT departments of hospitals are mainly responsible for the procurement, management, and subsequent maintenance of the system. The head of the IT department is the CIO of the hospital, and these CIOs are the main subjects of this research.

Technology Diffusion Model and Bass Modeling

Bass diffusion modeling was employed as one method to predict the progress of EMR adoption and analyze its characteristics. Diffusion theory is an essential branch of communication theory that has long attracted the attention of scholars in management, marketing, and other disciplines [11]. The Bass model has been widely used in the application and forecasting analysis of new products and technologies [12,13], including many medical-related technologies [14-16]. The Bass model has 9 key assumptions [13,16], most of which satisfy the scenarios of this study (eg, market potential of the new product remains constant over time, the geographic boundaries of the social system do not change over the diffusion process).

There are two important measures for the implementation of the Bass model [17]. The external influence coefficient is called the *innovation* effect, represented as the p-coefficient. It corresponds to the probability of using the products under the influence of public media or other external factors among users who have not used the product. The internal influence coefficient refers to the *imitation* effect and is expressed as the q-coefficient. This effect depicts the probability of the same users who would begin to use the product under the influence of peers who have already used the product [18]. The mathematical expression of the Bass model is shown in Figure 1, where M is the potential market, F(t) is the portion of M that have adopted by time t, p is the coefficient of innovation, and q is the coefficient of imitation.

Figure 1. Mathematical expression of the Bass model.

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}}$$

We conducted statistical analyses and forecasts using linear optimization in Excel for Mac 2011 (Microsoft Corp). The parameters of the Bass model were trained and estimated using SPSS Statistics software version 20 (IBM Corp). We used the method of least squares to determine the optimal values of q and p.

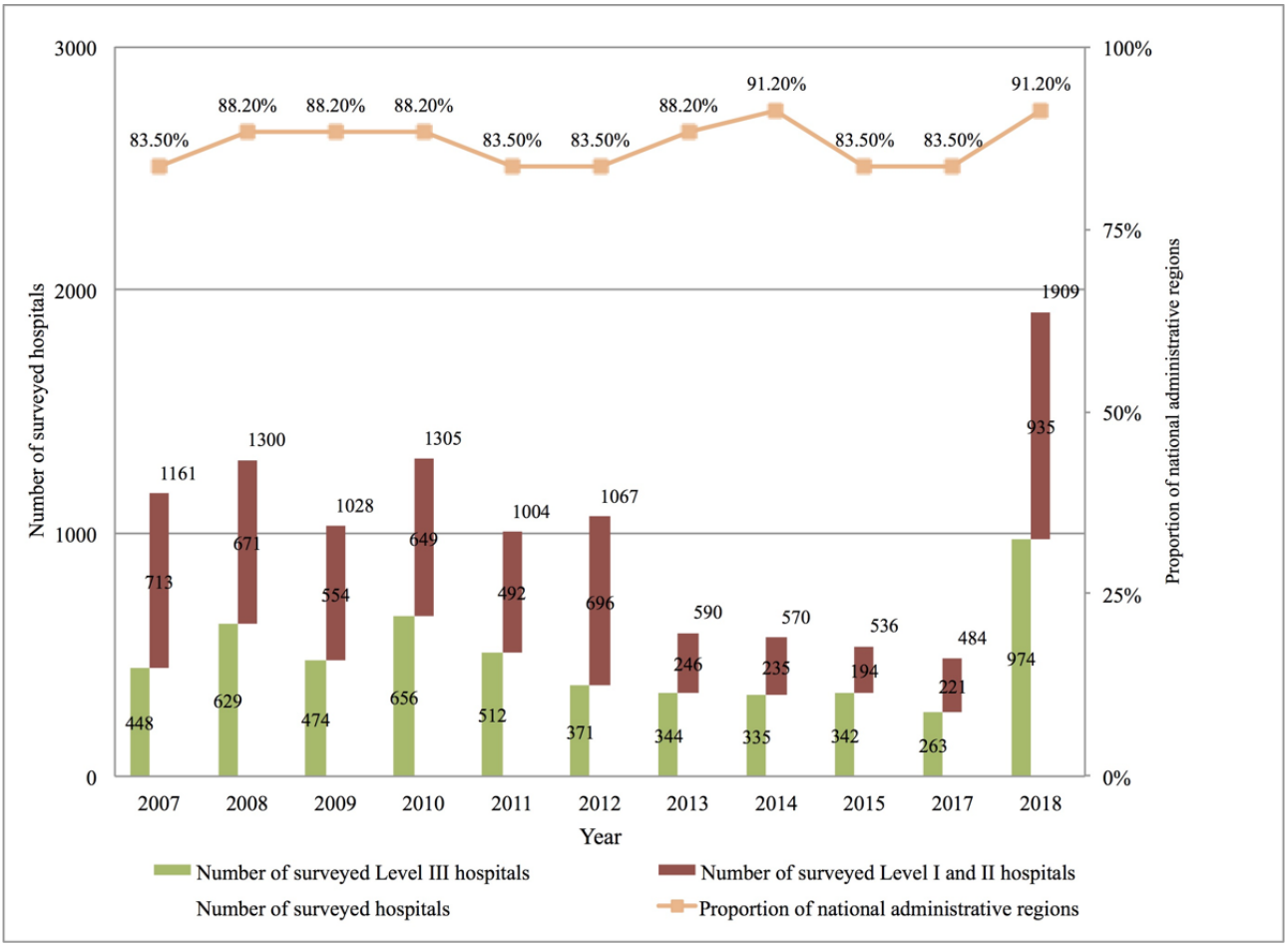
Results

Descriptive Analysis

Scale and Coverage of Research

The scale and regional coverage of the 2007-2018 CHIMA annual survey of hospital information systems are shown in Figure 2 below. In China, all hospitals are categorized by a government board into three levels: primary (roughly equivalent to community-based health centers in the United States), secondary (county- and municipal-level health care facilities), and level III (large, advanced general or specialty hospitals, often academic medical centers) [19]. In this study, hospitals were divided into two categories: level III and ≤ secondary.

Figure 2. China Hospital Information Management Association survey on hospital digitalization in China by hospital level, 2007 to 2018.



Hospital Health Information Technologies Human Resources: Quantity, Quality, and Work Stress

From 2007 to 2018, the shortage of human resources in China's hospital IT centers eased and the quality of personnel improved (Figures 3-6).

First, manpower allocation was 9.66 full-time equivalents (FTEs), on average, in 2018. At the same time, the average number of beds managed by each staff member in the hospital IT center decreased from 122 in 2007 to 93 in 2018, as shown in Figure 3. The proportion of IT centers in level III hospitals with 10 or more staff members increased from 27.44% in 2007 to 50.50% in 2018, as shown in Figure 4. However, compared with their US counterparts, the gap was still significant. According to the HIMSS annual survey data, as early as 2006, more than 80% of IT centers in US hospitals were staffed with more than 10 people [20].

Figure 3. Proportion of human resources in China's hospital information technology centers from 2007 to 2018. FTE: full-time equivalent; IT: information technology.

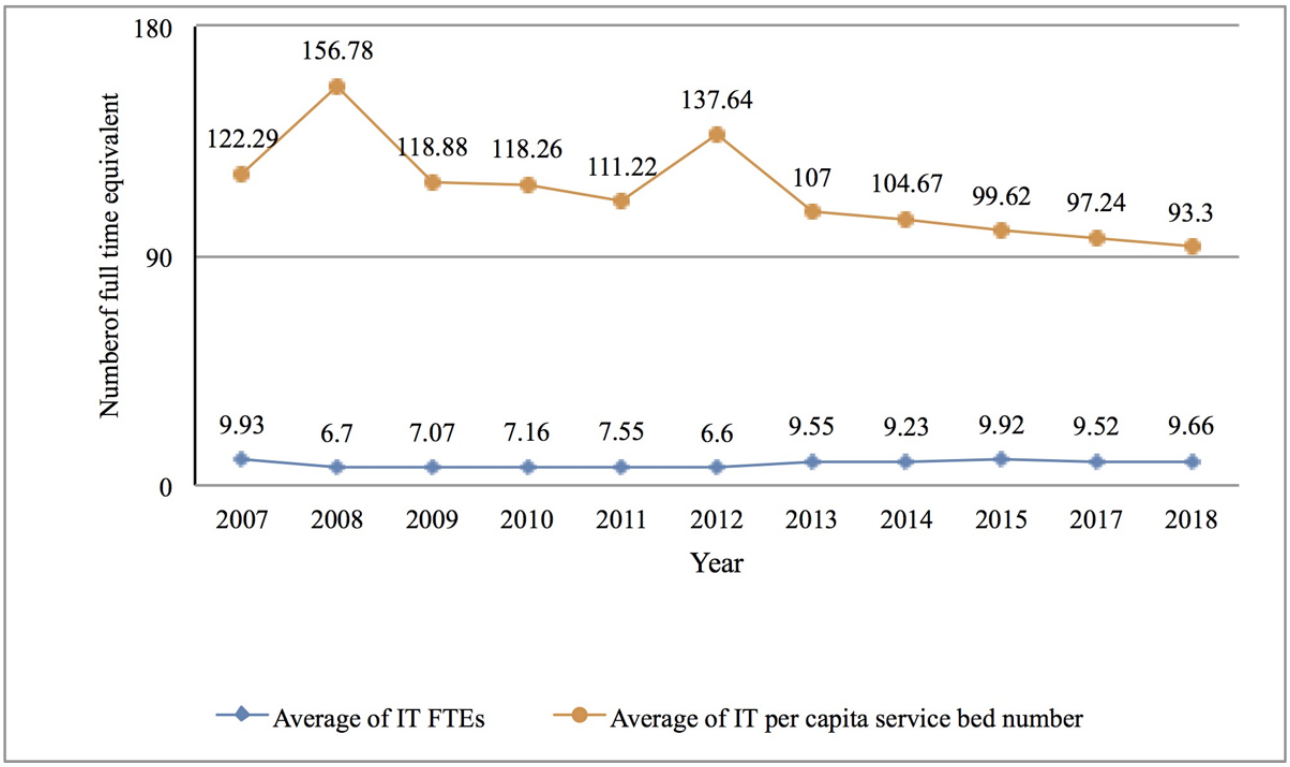
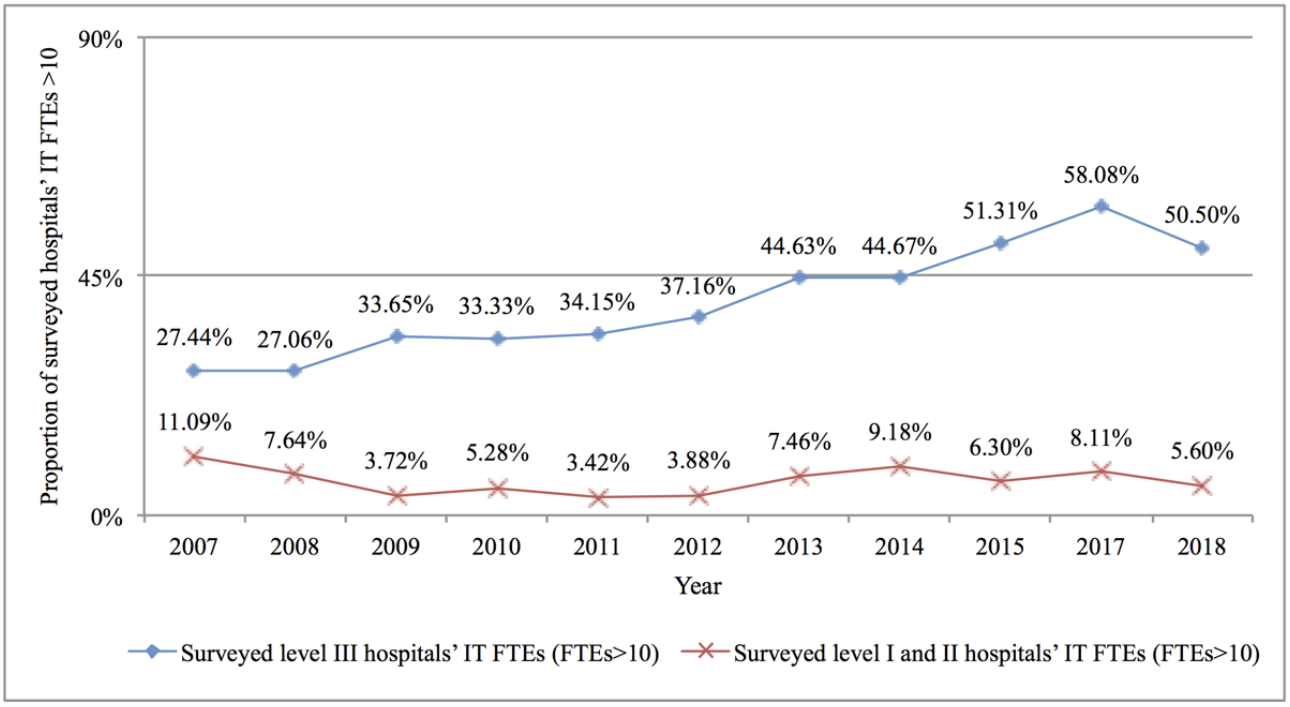


Figure 4. Information technology centers with 10 or more staff members in China's level III hospitals and hospitals below level III from 2007 to 2018. IT: information technology; FTE: full-time equivalent.



Second, the professional quality of CIOs in China's hospital IT centers also improved significantly. The proportion of hospital CIOs with a master's degree or above nearly tripled from 7.1% in 2007 to 25.61% in 2018. The proportion of CIOs with a master's degree in level III hospitals increased from 14.56% in 2007 to 42.17%, and the proportion of CIOs with a master's degree in level I and II hospitals increased from 2.08% to 6.31%, as shown in Figure 5. The proportion of CIOs with medical-related backgrounds in China's hospital IT centers was very low and even showed a downward trend, falling from 18.25% in 2007 to 11.37% in 2018, while computer majors became mainstream, rising from 41.95% in 2007 to 64.75% in 2018. As the counterpart discipline of HIT, medical informatics is in a marginally weak position among the background disciplines of CIOs in hospital IT centers, rising only from 2.24% in 2007 to 3.67% in 2018 (see Figure 6 for details).

Figure 5. Proportion of chief information officers with a master's degree or above in China's hospital information technology centers from 2007 to 2018. CIO: chief information officer.

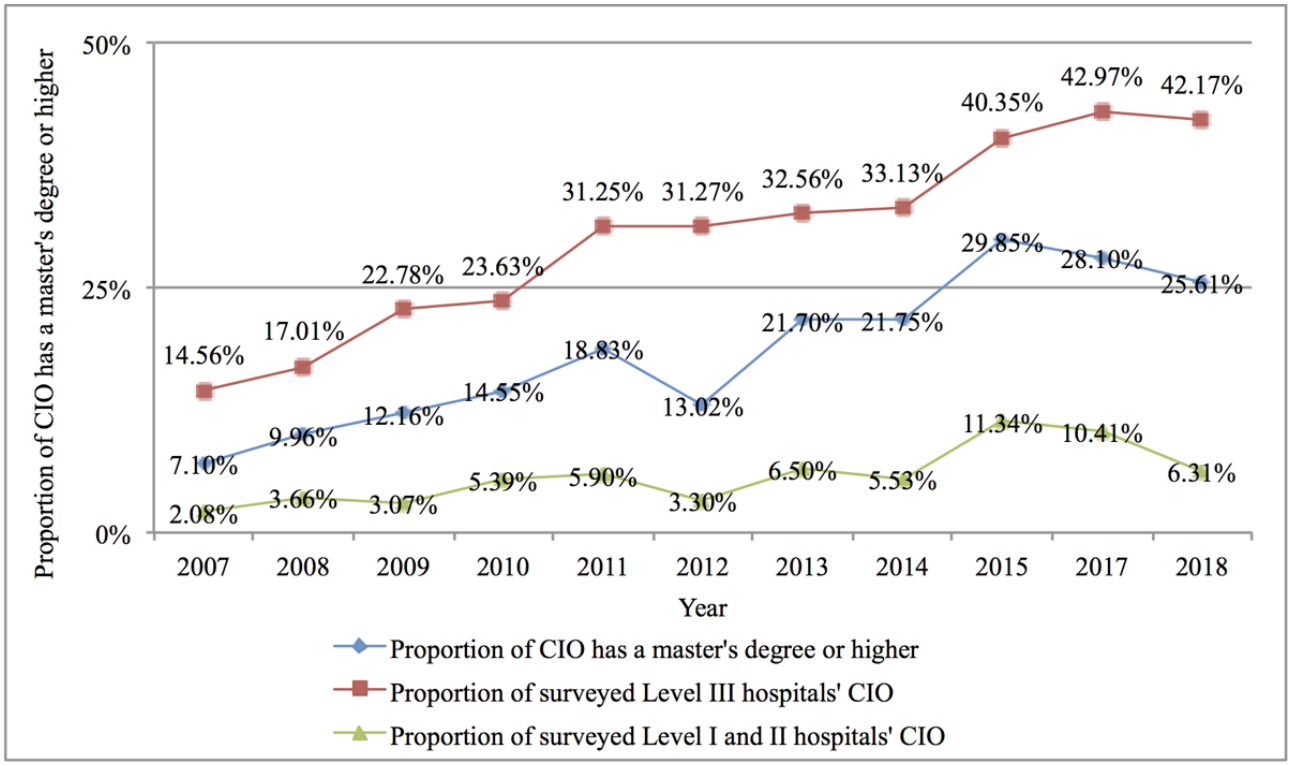
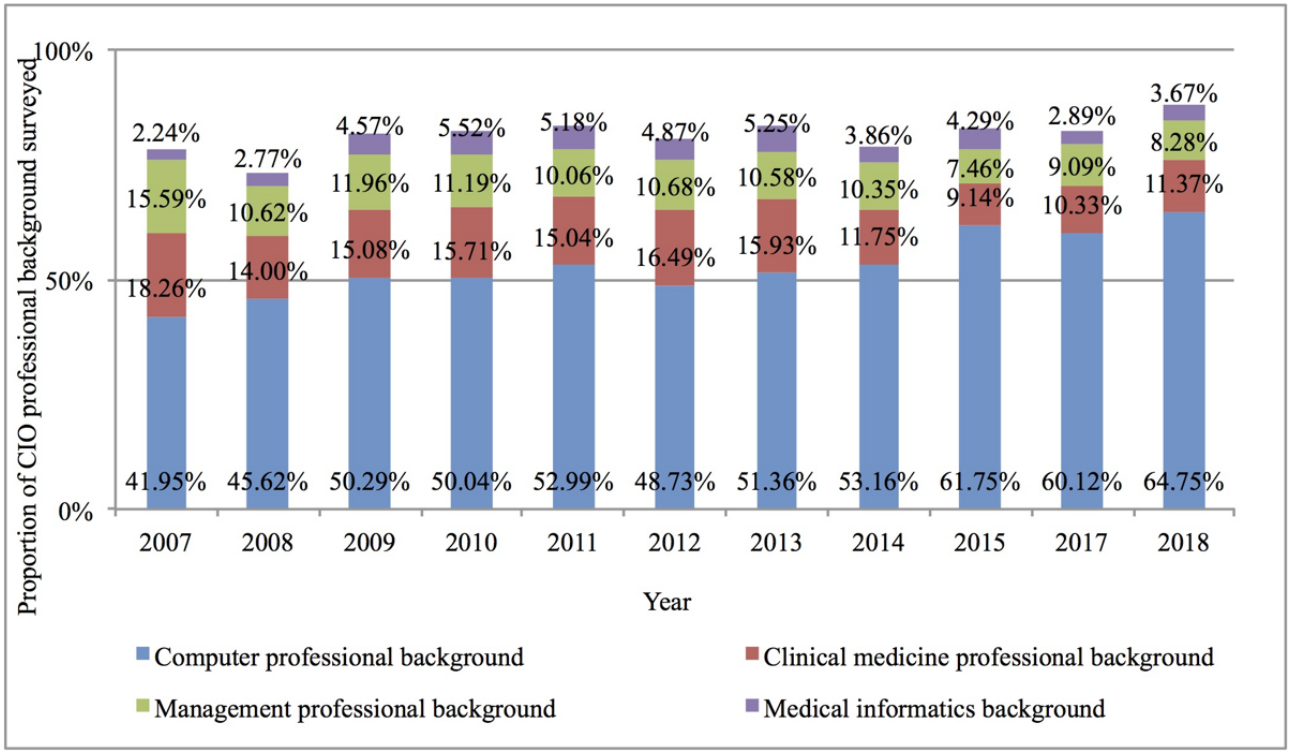


Figure 6. Academic background and composition of chief information officers in China's hospital information technology centers from 2007 to 2018. CIO: chief information officer.



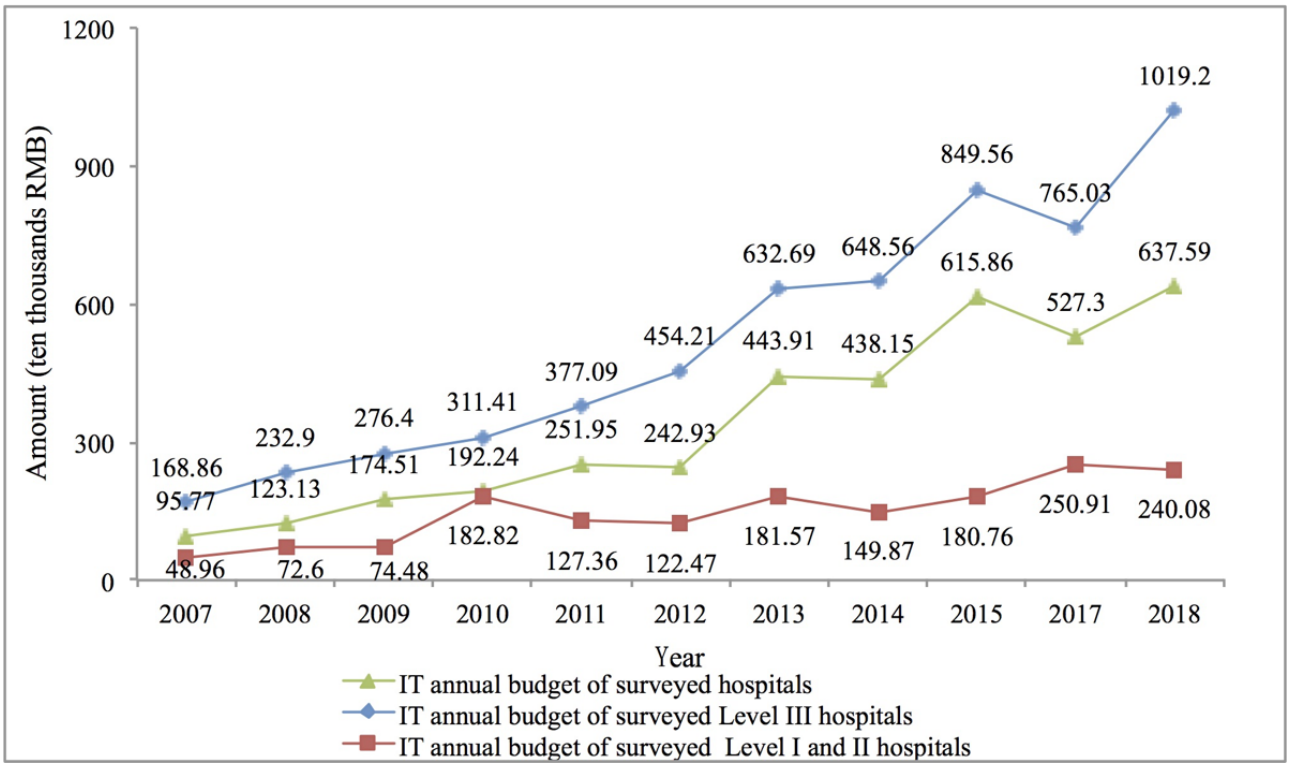
Hospital Health Information Technologies Investment

Stimulated and driven by the state's direct investment and relevant policies, the total direct investment by hospitals in HIT greatly increased.

First, the total investment in HIT rose from ¥957,700 (US $136,875) per year in 2007 to ¥6.376 million (US $0.91 million) per year in 2018, an increase of 5.66 times. The average annual HIT investment of level III hospitals increased from ¥1.689 million (US $0.24 million) per year to ¥10.192 million (US $1.46 million) per year, an increase of 5 times. The average

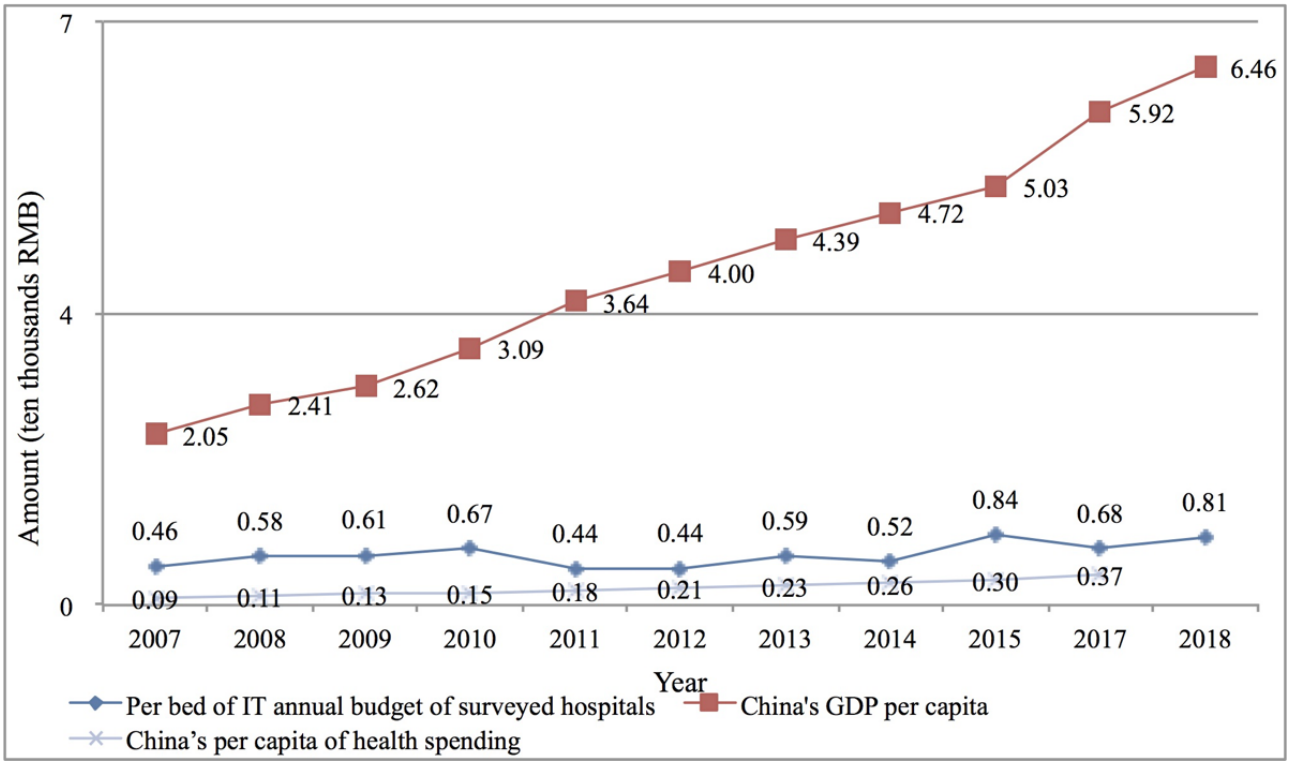
annual HIT investment of hospitals below level III increased from ¥489,600 (US $69,974) to ¥2.401 million (US $0.34 million) per year, an increase of nearly 4 times, as shown in Figure 7.

Figure 7. Health information technologies investment in Chinese hospitals from 2007 to 2018. IT: information technology.



Second, relative to China's fast-growing economy (per capita gross domestic product increased from ¥20,500 (US $2930) in 2007 to ¥64,600 (US $9233) in 2018, an increase of 2.15 times) and the rapid increase of medical expenses (per capita medical expenses increased from ¥900 (US $129) in 2007 to ¥3700 (US $528) in 2017, an increase of 4.28 times), the annual IT investment per bed increased insignificantly (only 76%) from ¥4600 (US $657) in 2007 to ¥8100 (US $1158) in 2018, as shown in Figure 8. However, due to the marginal cost of software and service products, the higher the base number of users, the larger the market, and the lower the cost of digitalization allocated to each single service object (bed). We believe that even considering the inflation factor, the connotation of the digitalization investment of ¥8100 (US $1158) per bed in 2018 was much greater than that of ¥4600 (US $657) in 2007.

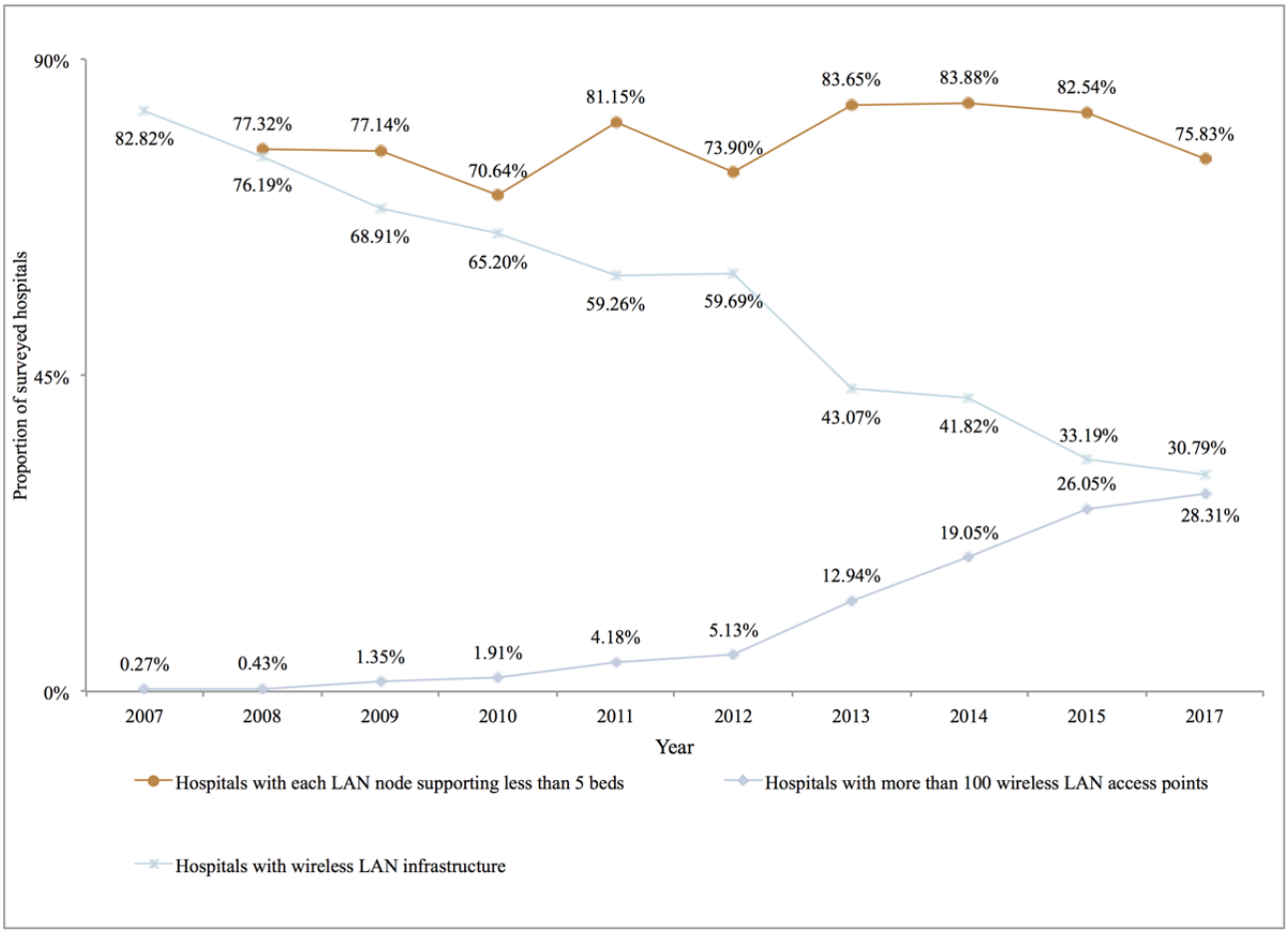
Figure 8. China's per capita gross domestic product, medical expenditure per capita, and information technology investment per hospital bed from 2007 to 2018. IT: information technology; GDP: gross domestic product. (Note: As of the date of submission, the per capita health spending data for China in 2018 has not been announced.).



Hospital Network Environment Support

The overall network infrastructure construction and configuration of Chinese hospitals have also been continuously improving. On one hand, in terms of traditional wired Ethernet local area network (LAN) construction, in 2017 about 75.83% of the sampled hospitals had achieved the goal of one wired LAN interface supporting 5 beds or fewer, which was basically the same as in 2008. On the other hand, in terms of wireless network infrastructure, about 69.21% of the sampled hospitals had launched wireless networks, compared with 17.18% in 2007. In addition, about 30.79% of the sampled hospitals that had launched wireless networks had more than 100 wireless network access hotspots in 2017, as shown in Figure 9.

Figure 9. Wired local area network and wireless network facility construction in Chinese hospitals from 2007 to 2017. LAN: local area network. (Note: Wireless network-related indicators were not included in the 2007 CHIMA Annual Survey; relevant indicators on hospital networks were no longer included in the 2018 CHIMA Annual Survey.).

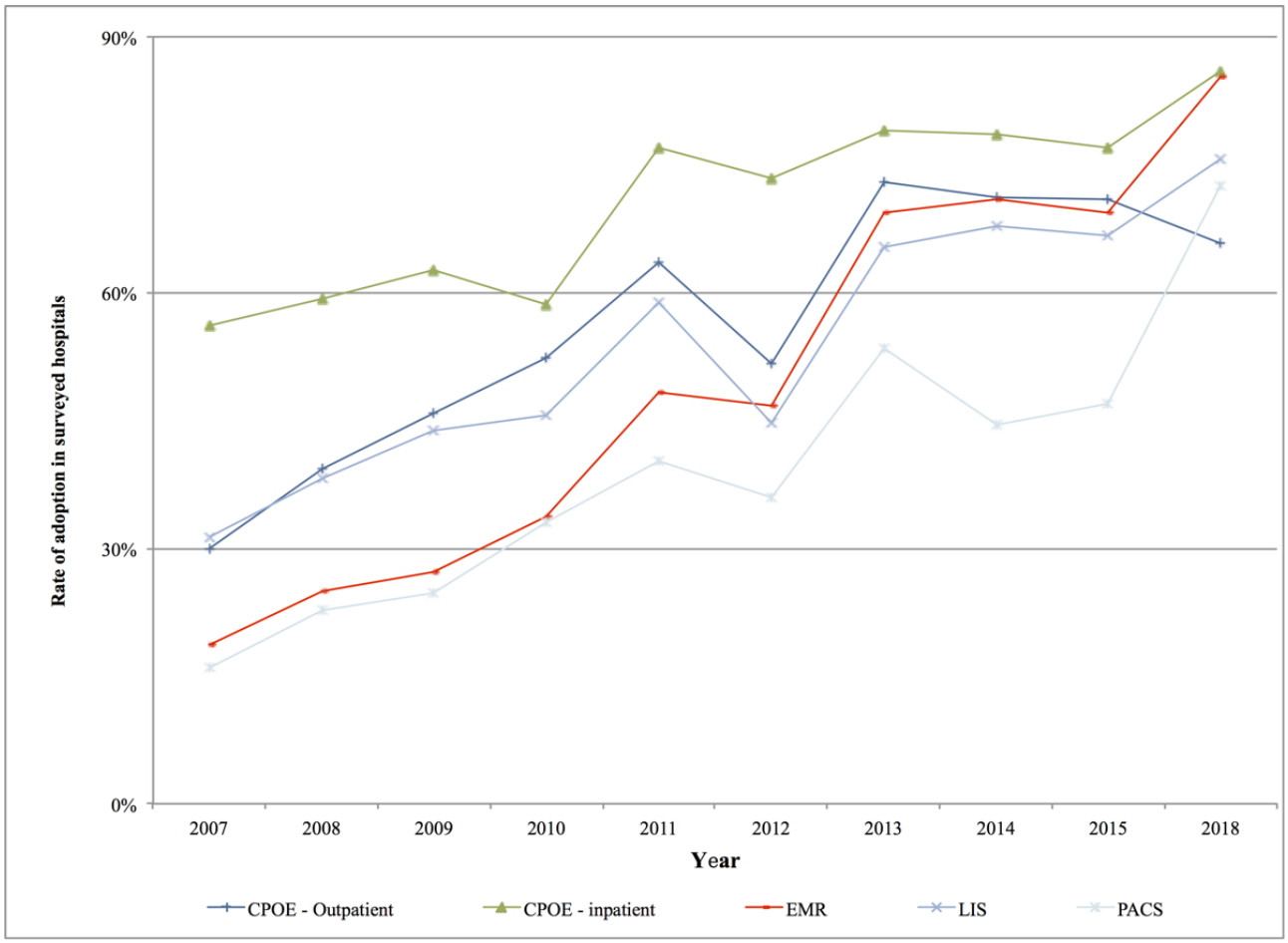


Implementation and Application of Clinical Information Systems (Including Electronic Medical Records) in Chinese Hospitals

CIS has been implemented in Chinese hospitals to a considerable extent. After more than 10 years of development, medical digitalization has been adopted as one of the "four beams and eight pillars" supporting China's health care reform, especially China's second health care reform, which began in 2010; a large amount of funds and resources have been invested, and a large number of policies have been promulgated for support and guidance [21]. Under this stimulus, from 2007 to 2018, the utilization rate of major CIS systems (including computerized prescriber order entry [CPOE], laboratory information systems [LIS], picture archiving and communication systems [PACS], and EMR) in sampled hospitals increased significantly.

CIS has been applied to a considerable extent, and the popularization rate of EMR exceeded the average level of its US counterparts in 2015 [22] (85.26% vs 83.8%) and the average level of its German counterparts in 2017 [23] (85.26% vs 68.4%). CPOE outpatient services rose from 30% in 2007 to 65.9% in 2018, CPOE inpatient services rose from 56.1% in 2007 to 85.9% in 2018, EMR rose from 18.6% in 2007 to 85.3% in 2018, LIS rose from 31.3% in 2007 to 75.7% in 2018, and PACS rose from 15.9% in 2007 to 72.5% in 2018, as shown in Figure 10. The construction and implementation of CIS including CPOE, EMR, LIS, and PACS in level III hospitals in China has developed vigorously and is maturing daily. In the 2018 survey, the utilization rates of CPOE, EMR, LIS, and PACS in the sampled hospitals all exceeded 65%. China's hospital digitalization focuses on the construction of a patient-centered clinical information system that directly serves medical personnel and provides strong support for health care reform.

Figure 10. Application and implementation of computerized prescriber order entry, electronic medical record, laboratory information system, and picture archiving and communication system in the sampled hospitals from 2007 to 2018. CPOE: computerized prescriber order entry; EMR: electronic medical record; LIS: laboratory information system; PACS: picture archiving and communication system. (Note: Due to a change in the leadership of CHIMA in 2016, the CHIMA Annual Survey was not launched, and survey data of 2016 and 2017 were not available for analysis.).



Bass Model Forecast Analysis: Development Trends of Electronic Medical Records in Chinese Hospitals

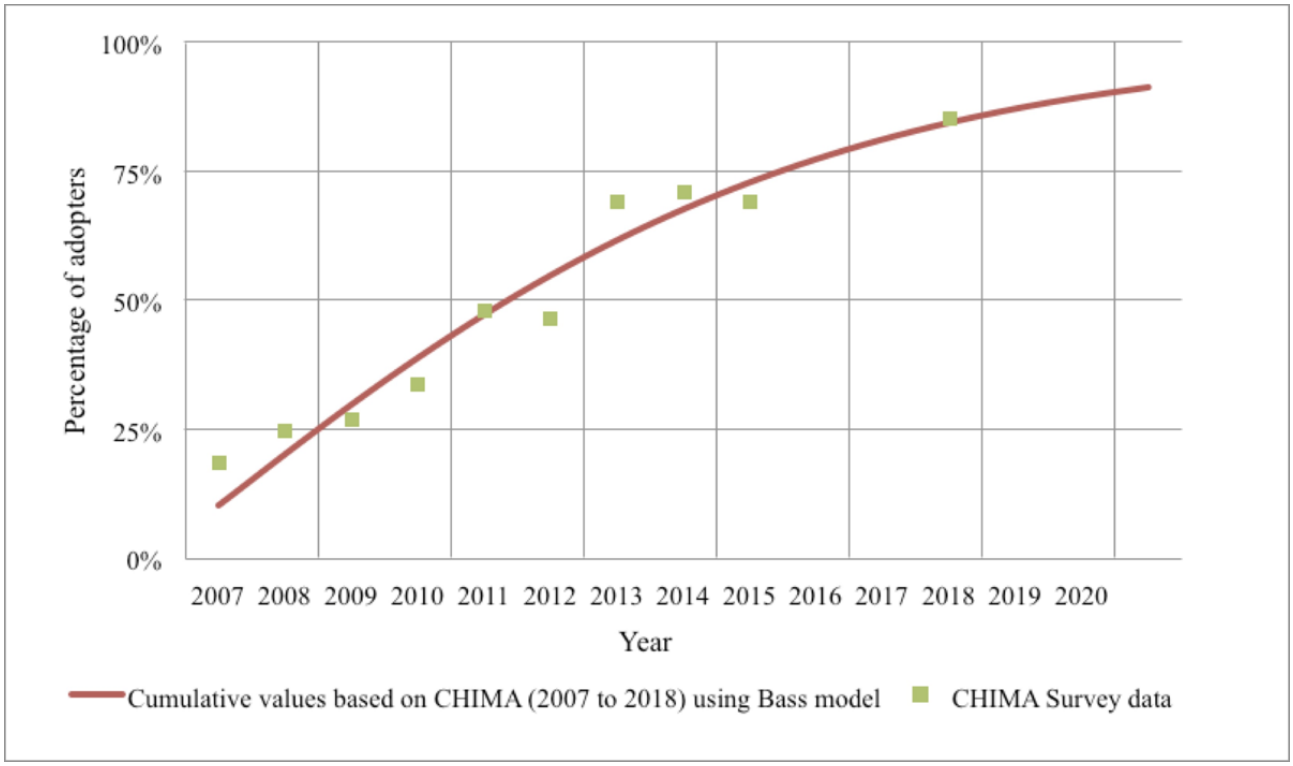
We estimated the p- and q-coefficients using the Bass model and linear optimization based on the CHIMA hospital adoption rate of EMR data from 2007-2018 (excluding 2016 and 2017; because of a leadership change in CHIMA in 2016, the 2016 annual survey was not launched). Table 1 describes the parameter estimation results in the final model, which indicates that the Bass model fit the CHIMA dataset well.

Figure 11 shows the fit of the EMR popularization data of Chinese hospitals from 2007-2018 (excluding 2016 and 2017), and assuming that there will be no major policy adjustments and technological upgrades in the future, the forecast of EMR popularization in hospitals by 2020 has the adoption rate of EMR expected to reach about 91.4%. Although great progress has been made, there is still a slight gap compared with US counterparts. According to research by Jha et al [24], the adoption of certified EHR systems among US nonfederal acute care hospitals in 2020 is expected to be close to 100%.

Table 1. The estimating parameters for Chinese hospitals' adoption rate of electronic medical records.

Model parameter	Estimated result
External motivation coefficient (p)	0.102
Internal motivation coefficient (q)	0.106
Motivation coefficient ratio (q/p)	1.039
R^2	.951

Figure 11. Electronic medical record adoption among Chinese hospitals using the China Hospital Information Management Association Annual Survey (figures for 2016, 2017, and 2020 are forecasted using the Bass mode). CHIMA: China Hospital Information Management Association.



On one hand, the fitted external coefficient (p=0.102) of Chinese hospitals' adoption rate of EMR is much larger than those of medical examination equipment popular in the United States, such as ultrasound images (p=0.000) and molybdenum target x-rays (p=0.000) [25]. On the other hand, it is relatively small compared with those of other consumer electronic products that provide information support, such as electronic calculators (p=0.143) and personal computers (p=0.121) [26]. At the same time, we found that for China's current medical system to join the regional medical care alliance, hospitals would need to invest significant manpower and material resources to build IT infrastructure and transform traditional paper-based medical processes and care, as well as the communication methods and business processes between doctors and patients, so as to provide medical services more effectively and efficiently. From this perspective, EMR may slow the spread of the universal technologies, such as personal computers and electronic calculators, that have been widely publicized. Despite this, the implementation and promotion of EMR have still been effectively advanced under the strong support and publicity of China's administrative supervision and public media.

On the other hand, compared with certain medical examination equipment, the internal motion coefficient (q) fitted in this study is relatively small, which indicates that the internal driving force of the hospitals themselves was relatively weak in this process. First, according to research by Sillup et al [25], the Bass model fitting parameter of ultrasound images in the United States was q=0.510 and the model fitting parameter of molybdenum target x-ray data was q=0.738, while in our study, q is only 0.106. Second, we believe that it cannot be immediately clear how much benefit hospitals can directly create for doctors and patients from the actual use of EMR. In many cases, the government took the lead, the public media and public opinions promoted it, and HIT technology, including EMR and CIS, was used to realize the artificial project of sharing medical resources in the hospital with information as the link rather than being a spontaneous product of the hospital.

Discussion

Summary

This study uses the data from the survey of medical digitalization construction conducted by CHIMA, a national industrial association in China, on 10,954 Chinese hospital CIOs from 2007 to 2018 to evaluate the progress of HIT in Chinese hospitals in terms of professional staffing, funding, infrastructure construction, and clinical system application. Here we discuss the US HIMSS annual survey exploring the difficulties and challenges encountered in the development of China's HIT.

Constraints on Health Information Technologies Human Resources

As of 2018, compared with their US counterparts, IT departments in Chinese hospitals were still short of IT human resources. The average allocation of human resources in the IT centers of the sampled Chinese hospitals was only 28% of that of their US counterparts in 2014 (9.66 FTEs vs 34 FTEs). We believe that this may further affect the development and deepening of subsequent HIT applications. In terms of the quantity of human resources, the survey results showed that hospital IT centers had an average of 9.66 FTEs in 2018, and the number of beds served by each IT staff member also dropped from 122 in 2007 to 93 in 2018. However, according to the annual survey of HIMSS in 2014, IT centers in the United States were equipped with an average of 34 FTEs, 3.5 times that of

their Chinese counterparts [27]. As early as 2006, more than 80% of the IT centers in US hospitals were equipped with 10 or more FTEs [20], while in China, by 2018, level III hospitals with 10 or more FTEs accounted for 50.5%, and hospitals classified as level II and below with 10 or more FTEs accounted for only 5.6%. Based on the results of the Information Statistics Center of the National Health Commission of the People's Republic of China in 2006, 10 to 30 HIT professionals were required for each level III hospital (600 beds or more), 6 to 15 for each level II hospital (300 to 600 beds), and 3 to 6 for each level I hospital (100 to 300 beds) [28].

Hospital information work such as system management; operation and maintenance; system and network security; content management; system integration and interface design; and hardware, network, and software maintenance is tedious and labor intensive, especially providing training for users of various levels and types of systems. Considering that many of the above services need to be provided on a 24/7 basis, it is an objective need and an inevitable trend for the development of hospital information systems to consume a large amount of human resources. We believe that, on one hand, the breadth and depth of HIT application in Chinese hospitals are still relatively low; on the other hand, policy makers and hospital managers do not fully understand that the safe and effective operation of information systems depends on the support of a large number of human resources.

The analysis of the highest degree of CIOs in hospitals indicates that the educational levels of information professionals working in hospitals in China had significantly improved; however, their distribution was not uniform. In 2018, 25.6% of CIOs had a master's degree or above, an increase of 18.51% compared with 2007; however, there was a significant difference between level III hospitals and hospitals below level III. Taking 2018 as an example, the proportion in the former was 35.8% higher than that in the latter. We believe that IT faces the urgent matter of cultivating interdisciplinary senior management talent who understand both medical care and IT technology. According to the survey results in 2018, more than 60% of the CIOs in China's hospital IT centers majored in computer information systems, while only 3.67% had a medical informatics background. Hospital CIOs demonstrated a relative lack of knowledge of hospital information management and medical informatics.

Unlike the cross-disciplinary definition of "using computer technology in the fields of health care and medical science" [29] in the United States, the medical informatics discipline in China is very young; however, it is gradually rising with the development of hospital digitalization in China on the basis of library science [30]. It was not formally established as an independent discipline until 2010, and at present, very few educational institutions in China have medical information research institutes or postgraduate programs (27 master's degree programs and 5 doctoral degree programs), and most of the current students are undergraduates who cannot meet the business needs of hospitals [31]. We suggest that reeducating experts interested in hospital digitalization in current leading positions in Chinese hospitals at all levels (systematically supplementing their knowledge of medical informatics based on International Medical Interpreters Association's training syllabus) and granting certificates to qualified personnel may be a shortcut to cultivating the required talent [32].

Hospital Health Information Technologies Investment Trends

HIT investment in a large number of hospitals classified as level II and below in China may be mainly driven by state investment, but their own investment willingness is not strong. After the previous health care reform, hospitals could only receive limited government financial subsidies and had to be self-financing [33]. Therefore, their financial strength was very limited. Beginning in 2010 (Figure 7), the average investment in HIT in Chinese hospitals increased rapidly, from ¥1.9224 million (US $0.27 million) in 2010 to ¥6.3759 million (US $0.91 million) in 2018, an increase of nearly 2.32 times. However, we found that the increase was extremely uneven (ie, after 2010, the HIT investment growth rate of level III hospitals was much higher than that of level II and below hospitals, and this imbalance may have caused new imbalances in medical resources).

According to an analysis of national HIT investment directions from 2010 to 2015 (Multimedia Appendix 1), the investment targets were mainly level II and below hospitals. According to the survey results, HIT investments in such hospitals in 2007, 2008, and 2009 were only ¥489,600 (US $69,974), ¥726,000 (US $103,760), and ¥744,800 (US $106,447), respectively. After 2010, HIT investment increased to ¥1.82 million (US $0.26 million), but subsequent growth was weak, with an increase of only ¥572,600 (US $81,836). During the same period, the increase for level III hospitals was ¥7.0779 million (US $1.01 million), which was 11.3 times the former. We suggest that the issue of how to raise the awareness of the majority of primary-level hospital leaders of the dividend that HIT brings to hospital development is one of the areas for which the National Health Commission should formulate relevant HIT development policies in the next stage.

Rapid Development of Electronic Medical Records in China and Difficulties in Recycled Use of Precipitated Data

The utilization rate of CIS represented by EMR in Chinese hospitals continued to increase. First, the EMR popularization rate of the sampled hospitals increased from 18.6% in 2007 to 85.3% in 2018, an increase of 3.6 times in 8 years, and the average EMR implementation rate of the sampled hospitals exceeded the average level of their US counterparts in 2015 [22] and their German counterparts in 2017 [23] (85.26% vs 83.8% vs 68.4%, respectively). Considering that as of 2017, the number of various medical institutions in China was more than 27,700, while that in the United States was more than 6300, the former close to 4.5 times of the latter, the growth rate was already considerable. Second, based on the Bass model fitting results of EMR utilization rate data from the sampled hospitals in the 2007-2015 CHIMA annual surveys, it is suggested that this growth was largely driven by external motivation coefficient effects (p-coefficient). That is, hospitals began to use EMR to a large extent under the influence of external administrative forces. The specific manifestation was p≈q (p=0.102, q=0.106),

which is consistent with the Chinese government's attitude and strategy toward HIT development. We believe that the development mode of China's medical industry, which accepts government instruction, uses unified planning of administrative intervention, and enables HIT to achieve leapfrog improvements in a short period of time, is one of the most important and unique contributions of China's HIT.

On the other hand, as CIS, represented by EMR, has gradually been built and put into use, it faces the challenge of how to carry out the secondary application of massive precipitated data in China. In the survey samples in 2018, the implementation rates of CPOE, EMR, LIS, and PACS all exceeded 65%. However, real-world clinical data from EMR and other CIS have not been widely used for secondary data research in China. The second health care reform in China established medical digitalization as an essential strategic development direction [34,35]. The reform also set the long-term goal of building and improving HIT, especially EMR software infrastructure in various hospitals. Based on the research feedback, the coverage of EMRs within hospitals reached 80% in 2018, which exceeds the coverage rate of hospitals in the United States in 2015. However, the interoperability, quality, and ease of use of EMR data are lacking.

In terms of interoperability, the various EMR systems used in different hospitals are incompatible with each other. There are currently more than 300 EMR software providers in China, all with their own proprietary technology structures and data standards. The hospitals have no initiative to exchange data despite the government establishment of some regional health information organizations (RHIOs). As of 2015, the proportion of hospitals participating in RHIOs in the sample had reached 50% [36]. Nevertheless, most of them are in the initial stages and are far from interoperability due to semantic problems.

Concerning the quality of information, EMR data in China is not informative. One study used Charmaz's grounded theory approach to perform a difference analysis of the medical questions and number of examination and treatment terminologies in the EMR corpus samples among 3 US hospitals and a Chinese hospital [37]. The study found that in certain types of medical records, the density of technical terms in Chinese EMRs was much lower than that in English EMRs. Chinese EMRs contained only half the amount of technical terms compared to US EMRs, indicating that the latter is more professional. We believe that this may be due to the more complicated and rigorous legal environment in the United States, where more complete and comprehensive examinations and discussions with patients are required to prevent medical disputes.

Regarding ease of use, there are large discrepancies and gaps between EMR data in China and the United States. This indirectly leads to problems of integrity and accuracy in China's EMR data. Previous research used the US Stage 2 Meaningful Use objectives to evaluate usability of EMR data from the two best Chinese teaching hospitals affiliated with Peking University Medical School (Peking University First Hospital and Beijing Cancer Hospital) [38]. They found that only 50% of the Meaningful Use targets were supported in the EMRs of Chinese hospitals. Moreover, the Chinese hospitals still used many paper forms to augment the clinical work despite the establishment of EMRs, resulting in a considerable loss of clinical information beyond the EMR system. The ease of use of EMRs at Peking University First Hospital and Beijing Cancer Hospital [39] was examined based on the standard of the Unified Framework For EHR Usability [40], and a total of 85 problems in usability relevant to clinical tasks were found, some of which may even seriously affect the quality and safety of medical services.

Limitations

This study is based on self-reported questionnaire survey results from 2007-2018 regarding investment in HIT funds, staffing and training, investment in funds, construction and implementation of applied technologies, and difficulties encountered in the processes of Chinese hospitals. The data have not been independently verified. Therefore, such an analysis is subject to the potential confounding factors of data bias. In addition, we did not use a multivariate model to evaluate the independence of different factors (such as hospital level, hospital type, and economic development level in the region of the hospital). Although we only limit the inference to our own samples, these analyses are still valuable because these data spanning 12 years are the only data on the development trend of HIT in China collected by China's national industrial association that can be quantitatively analyzed.

In addition, the absence of feedback on data offset will affect the survey results. For example, hospitals with high HIT application levels are more likely to give feedback. However, the feedback providers of this survey should be representative of the true level of HIT application in Chinese hospitals to some extent, especially for the level III hospitals in China, which have an average coverage rate of 34.44% over 12 years.

Conclusions

China's unique institutional model may have distinct advantages in achieving the goals of health care reform. In this case, the Chinese government used a top-down, top-level design mode and took HIT development as an important technical support and starting point to support health care reform through policies, systems, funds, and other comprehensive methods. According to the survey results of the CHIMA annual survey of hospital information systems, with about only one-fifth of the required funding and one-fourth of the required human resources funding per hospital IT FTE as compared with the US HITECH project, China's EMR coverage in 2018 exceeded the average level of its US counterpart in 2015 and the average level of its German counterpart in 2017. Fitting results based on the Bass model suggest that it is expected that 91% of hospitals in China will use EMR by 2020. All signs show that the Chinese government is gradually approaching and realizing the phased goals set in the second health care reform launched in 2010: integrating medical resources, improving medical care popularization, reducing medical costs, and improving medical care quality.

Acknowledgments

We thank Elsevier language editing. This work was supported by the National Natural Science Foundation of China (grant numbers 81771937, 81871455).

Conflicts of Interest

None declared.

Multimedia Appendix 1
National health information technologies investment directions from 2010 to 2015.
[DOCX File , 20 KB - medinform_v8i2e17006_app1.docx]

References

1. Kim Y, Jung K, Park Y, Shin D, Cho S, Yoon D, et al. Rate of electronic health record adoption in South Korea: a nation-wide survey. Int J Med Inform 2017 May;101:100-107. [doi: 10.1016/j.ijmedinf.2017.02.009] [Medline: 28347440]
2. Mennemeyer ST, Menachemi N, Rahurkar S, Ford EW. Impact of the HITECH Act on physicians' adoption of electronic health records. J Am Med Inform Assoc 2016 Mar;23(2):375-379. [doi: 10.1093/jamia/ocv103] [Medline: 26228764]
3. Gold M, McLaughlin C. Assessing HITECH implementation and lessons: 5 years Later. Milbank Q 2016 Sep;94(3):654-687 [FREE Full text] [doi: 10.1111/1468-0009.12214] [Medline: 27620687]
4. Sheikh A, Jha A, Cresswell K, Greaves F, Bates DW. Adoption of electronic health records in UK hospitals: lessons from the USA. Lancet 2014 Jul 05;384(9937):8-9. [doi: 10.1016/S0140-6736(14)61099-0] [Medline: 24998803]
5. Lei J, Wen D, Zhang X, Li J, Lan H, Meng Q, et al. Enabling health reform through regional health information exchange: a model study from China. J Healthc Eng 2017(2017). [doi: 10.1155/2017/1053403] [Medline: 29068625]
6. Zhu C. Healthy China 2020 Strategic Research Report. 1st Edition. Philadelphia: Elsevier; 2015.
7. State Council, the People's Republic of China. 2018 Aug 15. [Directive on further promoting the informatization construction of medical institutions with electronic medical records as the core] URL: http://www.nhc.gov.cn/yzygj/s7659/201808/a924c197326440cdaaa0e563f5b111c2.shtml [accessed 2020-01-03]
8. Kruse CS, Beane A. Health information technology continues to show positive effect on medical outcomes: systematic review. J Med Internet Res 2018 Feb 05;20(2):e41 [FREE Full text] [doi: 10.2196/jmir.8793] [Medline: 29402759]
9. Adler-Milstein J, DesRoches CM, Furukawa MF, Worzala C, Charles D, Kralovec P, et al. More than half of US hospitals have at least a basic EHR, but stage 2 criteria remain challenging for most. Health Aff (Millwood) 2014 Sep;33(9):1664-1671. [doi: 10.1377/hlthaff.2014.0453] [Medline: 25104826]
10. China Hospital Information Management Association. URL: http://www.chima.org.cn/ [accessed 2019-12-16]
11. Bass FM. A new product growth for model consumer durables. Manag Sci 2004 Dec;50(12_supplement):1825-1832. [doi: 10.1287/mnsc.1040.0264]
12. Van den Bulte C. Want to know how diffusion speed varies across countries and products? Try using a Bass model. PDMA visions 2002;26(4):12-15. [doi: 10.1002/9781444316568.wiem05030]
13. Sood A, James GM, Tellis GJ, Zhu J. Predicting the path of technological innovation: SAW vs. Moore, Bass, Gompertz, and Kryder. Market Sci 2012 Nov;31(6):964-979. [doi: 10.1287/mksc.1120.0739]
14. Kharrazi H, Gonzalez CP, Lowe KB, Huerta TR, Ford EW. Forecasting the maturation of electronic health record functions among US hospitals: retrospective analysis and predictive model. J Med Internet Res 2018 Aug 07;20(8):e10458 [FREE Full text] [doi: 10.2196/10458] [Medline: 30087090]
15. Ford EW, Hesse BW, Huerta TR. Personal health record use in the United States: forecasting future adoption levels. J Med Internet Res 2016 Mar 30;18(3):e73 [FREE Full text] [doi: 10.2196/jmir.4973] [Medline: 27030105]
16. Norton JA, Bass FM. A diffusion theory model of adoption and substitution for successive generations of high-technology products. Manag Sci 1987 Sep;33(9):1069-1086. [doi: 10.1287/mnsc.33.9.1069]
17. Mahajan V, Muller E, Bass FM. New product diffusion models in marketing: a review and directions for research. J Marketing 2018 Nov 28;54(1):1-26. [doi: 10.1177/002224299005400101]
18. Grimm SE, Stevens JW, Dixon S. Estimating future health technology diffusion using expert beliefs calibrated to an established diffusion model. Value Health 2018 Aug;21(8):944-950 [FREE Full text] [doi: 10.1016/j.jval.2018.01.010] [Medline: 30098672]
19. Lei J, Guan P, Gao K, Lu X, Chen Y, Li Y, et al. Characteristics of health IT outage and suggested risk management strategies: an analysis of historical incident reports in China. Int J Med Inform 2014 Feb;83(2):122-130. [doi: 10.1016/j.ijmedinf.2013.10.006] [Medline: 24246272]
20. PricewaterhouseCoopers LLP. California HealthCare Foundation. 2007 Jun. The Financial Health of California Hospitals URL: https://www.issuelab.org/resources/9288/9288.pdf?download=true [accessed 2020-01-03]
21. Lei J, Meng Q, Li Y, Liang M, Zheng K. The evolution of medical informatics in China: a retrospective study and lessons learned. Int J Med Inform 2016 Aug;92:8-14. [doi: 10.1016/j.ijmedinf.2016.04.011] [Medline: 27318067]

22. Charles D, Gabriel M, Searcy T. Adoption of electronic health record systems among US non-federal acute care hospitals: 2008-2014. Washington: Office of the National Coordinator for Health Information Technology; 2015 Apr. URL: https://www.healthit.gov/sites/default/files/data-brief/2014HospitalAdoptionDataBrief.pdf [accessed 2019-12-16]
23. Esdar M, Hüsers J, Weiß J, Rauch J, Hübner U. Diffusion dynamics of electronic health records: a longitudinal observational study comparing data from hospitals in Germany and the United States. Int J Med Informatics 2019 Nov;131:103952 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.103952]
24. Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital. J Am Med Inform Assoc 2017 Nov 01;24(6):1142-1148. [doi: 10.1093/jamia/ocx080] [Medline: 29016973]
25. Sillup GP. Forecasting the adoption of new medical technology using the Bass model. J Health Care Mark 1992 Dec;12(4):42-51. [Medline: 10123584]
26. Ford EW, Menachemi N, Phillips MT. Predicting the adoption of electronic health records by physicians: when will health care be paperless? J Am Med Inform Assoc 2006;13(1):106-112 [FREE Full text] [doi: 10.1197/jamia.M1913] [Medline: 16221936]
27. Healthcare Information and Management Systems Society. 2014 Feb 24. 25th Annual HIMSS Leadership Survey—Final Report: Healthcare CIO URL: http://s3.amazonaws.com/rdcms-himss/files/production/public/FileDownloads/2014-HIMSS-Leadership-Survey.pdf [accessed 2020-01-03]
28. Luo S, Zhang K, Li B. Medical informatics in China: healthcare IT trends, academic and research developments. Yearb Med Inform 2010:109-115. [Medline: 20938582]
29. Musen M, Bemmel H. Handbook of Medical Informaties. Houten: Springer; 2002.
30. Liang J, Wei K, Meng Q, Chen Z, Zhang J, Lei J. The gap in medical informatics and continuing education between the United States and China: a comparison of conferences in 2016. J Med Internet Res 2017 Dec 21;19(6):e224 [FREE Full text] [doi: 10.2196/jmir.8014] [Medline: 28637638]
31. Liang J, Wei K, Meng Q, Chen Z, Zhang J, Lei J. Development of medical informatics in China over the past 30 years from a conference perspective and a Sino-American comparison. PeerJ 2017;5:e4082 [FREE Full text] [doi: 10.7717/peerj.4082] [Medline: 29177118]
32. Lau F. Distributed health informatics graduate education for working professionals. Int J Med Inform 2007;76(5-6):344-350. [doi: 10.1016/j.ijmedinf.2007.01.008] [Medline: 17307030]
33. Editorial. Chinese doctors are under threat. Lancet 2010 Aug 28;376(9742):657. [doi: 10.1016/S0140-6736(10)61315-3] [Medline: 20801385]
34. Deng H, Wang J, Liu X, Liu B, Lei J. Evaluating the outcomes of medical informatics development as a discipline in China: a publication perspective. Comput Methods Programs Biomed 2018 Oct;164:75-85. [doi: 10.1016/j.cmpb.2018.07.001] [Medline: 30195433]
35. Jia Y, Wang W, Liang J, Liu L, Chen Z, Zhang J, et al. Trends and characteristics of global medical informatics conferences from 2007 to 2017: a bibliometric comparison of conference publications from Chinese, American, European and the Global Conferences. Comput Methods Programs Biomed 2018 Nov;166:19-32. [doi: 10.1016/j.cmpb.2018.08.017] [Medline: 30415715]
36. Liang J, Zheng X, Chen Z, Dai S, Xu J, Ye H, et al. The experience and challenges of healthcare-reform-driven medical consortia and Regional Health Information Technologies in China: a longitudinal study. Int J Med Inform 2019 Nov;131:103954. [doi: 10.1016/j.ijmedinf.2019.103954] [Medline: 31513943]
37. Wu Y, Lei J, Wei W, Tang B, Denny JC, Rosenbloom ST, et al. Analyzing differences between chinese and english clinical text: a cross-institution comparison of discharge summaries in two languages. Stud Health Technol Inform 2013;192:662-666 [FREE Full text] [Medline: 23920639]
38. Lei J, Sockolow P, Guan P, Meng Q, Zhang J. A comparison of electronic health records at two major Peking University Hospitals in China to United States meaningful use objectives. BMC Med Inform Decis Mak 2013 Aug 28;13:96 [FREE Full text] [doi: 10.1186/1472-6947-13-96] [Medline: 23984797]
39. Xu L, Wen D, Zhang X, Lei J. Assessing and comparing the usability of Chinese EHRs used in two Peking University hospitals to EHRs used in the US: a method of RUA. Int J Med Inform 2016 May;89:32-42. [doi: 10.1016/j.ijmedinf.2016.02.008] [Medline: 26980357]
40. Zhang J, Walji MF. TURF: toward a unified framework of EHR usability. J Biomed Inform 2011 Dec;44(6):1056-1067 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.005] [Medline: 21867774]

Abbreviations

CHIMA: China Hospital Information Management Association
CIS: clinical information system
CIO: chief information officer
CPOE: computerized prescriber order entry
EHR: electronic health record

EMR: electronic medical record
FTE: full-time equivalent
HIMSS: Healthcare Information and Management Systems Society
HIT: health information technologies
HITECH: Health Information Technology for Economic and Clinical Health
IT: information technology
LAN: local area network
LIS: laboratory information system
PACS: picture archiving and communication system
RHIO: regional health information organization

Edited by Z Huang, G Eysenbach; submitted 12.11.19; peer-reviewed by C Liang, D Liu; comments to author 29.11.19; revised version received 10.12.19; accepted 11.12.19; published 10.02.20.

Please cite as:
Liang J, Li Y, Zhang Z, Shen D, Xu J, Yu G, Dai S, Ge F, Lei J
Evaluating the Applications of Health Information Technologies in China During the Past 11 Years: Consecutive Survey Data Analysis
JMIR Med Inform 2020;8(2):e17006
URL: https://medinform.jmir.org/2020/2/e17006
doi:10.2196/17006
PMID:32039815

Original Paper

Intellectual Structure and Evolutionary Trends of Precision Medicine Research: Coword Analysis

Xiaoguang Lyu[1], PhD; Jiming Hu[2,3], PhD; Weiguo Dong[1], PhD; Xin Xu[4], BA

[1]The Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan, China

[2]School of Information Management, Wuhan University, Wuhan, China

[3]Center for the Study of Information Resources, Wuhan University, Wuhan, China

[4]The Intensive Care Unit of Coronary Heart Disease, Renmin Hospital of Wuhan University, Wuhan, China

Corresponding Author:
Jiming Hu, PhD
School of Information Management
Wuhan University
Wuchang
Wuhan, 430072
China
Phone: 86 18995621959
Email: hujiming@whu.edu.cn

Abstract

Background: Precision medicine (PM) is playing a more and more important role in clinical practice. In recent years, the scale of PM research has been growing rapidly. Many reviews have been published to facilitate a better understanding of the status of PM research. However, there is still a lack of research on the intellectual structure in terms of topics.

Objective: This study aimed to identify the intellectual structure and evolutionary trends of PM research through the application of various social network analysis and visualization methods.

Methods: The bibliographies of papers published between 2009 and 2018 were extracted from the Web of Science database. Based on the statistics of keywords in the papers, a coword network was generated and used to calculate network indicators of both the entire network and local networks. Communities were then detected to identify subdirections of PM research. Topological maps of networks, including networks between communities and within each community, were drawn to reveal the correlation structure. An evolutionary graph and a strategic graph were finally produced to reveal research venation and trends in discipline communities.

Results: The results showed that PM research involves extensive themes and, overall, is not balanced. A minority of themes with a high frequency and network indicators, such as Biomarkers, Genomics, Cancer, Therapy, Genetics, Drug, Target Therapy, Pharmacogenomics, Pharmacogenetics, and Molecular, can be considered the core areas of PM research. However, there were five balanced theme directions with distinguished status and tendencies: Cancer, Biomarkers, Genomics, Drug, and Therapy. These were shown to be the main branches that were both focused and well developed. Therapy, though, was shown to be isolated and undeveloped.

Conclusions: The hotspots, structures, evolutions, and development trends of PM research in the past ten years were revealed using social network analysis and visualization. In general, PM research is unbalanced, but its subdirections are balanced. The clear evolutionary and developmental trend indicates that PM research has matured in recent years. The implications of this study involving PM research will provide reasonable and effective support for researchers, funders, policymakers, and clinicians.

(JMIR Med Inform 2020;8(2):e11287) doi:10.2196/11287

KEYWORDS

precision medicine; topics distribution; correlation structure; evolution patterns; coword analysis

Introduction

Background

Precision medicine (PM), also called personalized medicine, is a new medical model aimed at providing precise diagnosis, therapy, prognosis prediction, and prevention strategies based on information in a patient's genes, proteins, and their environment [1]. The scientific basis of PM is molecular pathological epidemiology, and it aims to identify the relationship between biomarkers, the drug response, and outcome in disease [2,3]. During the Human Genome Project, it took "one dollar one bp" and 13 years to complete the sequencing of the whole genome. Owing to breakthroughs in techniques and lower prices, effective, high-throughput, and accurate sequencing can be applied to map genomics, metabolomics, microbiomics, and proteomics, which has led to the discovery of increasingly more causative biomarkers [4-7]. However, clinicians currently use clinical trials and pilot studies to assess the relationship between biomarkers and diseases.

Great progress has been made in personalized treatment in the field of oncology. According to a meta-analysis of phase II clinical trials, a personalized treatment strategy across malignancies yields a better outcome and lower likelihood of death than nonpersonalized targeted therapies [8]. Thus, it can be expected that PM will use new knowledge, including the integration of clinical medicine, pathology, epidemiology, and omics, to provide better therapies for patients.

Owing to the potential importance of PM, a few leading experts reviewed this new medical approach in regards to its relevant history, clinical applications, and any interdisciplinary research associated with PM, such as bioinformatics, artificial intelligence, and big data [9-11]. It is logical to assess the status of the subfields or branches. However, there are still some limitations regarding the review themes; specifically, the overall structure and characteristics of PM research have not been mapped, the relationship between the subfields has not been revealed, and the predictions regarding PM in those reviews were not made based on an accurate quantitative analysis.

Coword analysis is a bibliometric method used to identify relationships between subfields within research areas and to measure the strength of the relationships [12,13]. According to the co-occurrence correlation, the keywords can be classified into clusters and displayed as network maps. Some other indices, such as density and centrality, can be used to evaluate the shape of the maps. By comparing the network maps of different periods, the dynamic evolution of one research area can be clearly displayed. Owing to the characteristics of "quantitative" and "zoom" in coword analysis, scientists can uncover the links within a subfield, obtain the overall structure of networks according to simplified graphs, and focus on one certain subarea to obtain more information [13].

Coword analysis has been widely used to illustrate the intellectual structure and developmental status of research areas [14-16]. Our study applied this method to explore the overall research structure, correlation among themes, and entire set of evolutionary trends in the field of PM. Our results may help scientists and clinicians better understand its developmental characteristics and even yield new insights on breakthroughs.

Literature Review

PM is a new medical approach that classifies patients into different groups related to their diagnosis, treatment, and prevention based on individual gene, protein, or environmental information. It is noteworthy that the terms "precision medicine," "personalized medicine," "stratified medicine," and "P4 [predictive, preventative, personalized, and participatory] medicine" are still interchangeably used by some organizations and scientists [17,18]. In the period after the Human Genome Project, considerably more effort was put into exploring the relationship between genomic information and patient care [19]. In 2015, the Precision Medicine Initiative was launched by Barack Obama, then the president of the United States, indicating the beginning of a new medical age [20].

Every person has polymorphisms in their DNA, RNA, and proteins, as well as methylation. Recent scientific methods have enabled the analysis of biomarkers using omics techniques, including genomics, epigenomics, transcriptomics, proteomics, metabolomics, microbiome analysis, and immunomics. However, pathologists, epidemiologists, and clinicians have also made many contributions to the discovery of links between biomarkers and clinical features [21]. Advances in the PM model have played an important role in disease diagnosis, treatment, and prevention. Cancer treatment is the field in which PM originated and it has seen the most mature use of PM, such as when cancer genomics were successfully applied in personalized medicine. With the deep awareness of genetic variations of tumors, treatment strategy can be tailored to the group of cancer patients with the same genotype. For example, the human epidermal growth factor receptor 2 (HER2) gene is amplified and overexpressed in 25-30% of breast cancers. According to evidence from clinical trials, trastuzumab, a targeted therapy drug, increases the clinical benefit of first-line chemotherapy in metastatic breast cancer that overexpresses HER2 [22].

PM also plays an important role in disease treatment and prevention. DNA information from individual phenotyping will lead to more effective and accurate treatment and prevention. For example, the high risk in women for developing breast cancer is strongly correlated with mutations in *BRCA1* or *BRCA2* [23]. Clinicians can make better decisions regarding prevention for patients carrying these genetic mutations. Pharmacologists and genomic scientists have also provided many contributions to the assessment of genetic variations that affect drug discovery and clinical pharmacology [24]. Considering information on personal phenotypes, physicians can provide reasonable drug prescriptions that represent targeted therapies and are more cost-effective, but also have fewer side effects [25]. Furthermore, PM has changed clinical trials. Among the minority of clinical trials involving PM, the proportion of trials on adult cancers in the United States that require a genomic alteration for enrollment has increased substantially over the past several years [26]. Finally, PM will yield some challenges in the field of ethics, patient privacy, and refurbishment policies, which will require more attention from scientists and policymakers in the field [27,28].

Previous Efforts

With the pace of PM research rapidly increasing, a large number of studies have been performed from different perspectives. Many reviews have been published to facilitate a better understanding of the status of PM research, as well as clarifying the concept, history, clinical application, ethical concerns, and technological challenges. The efforts listed above have helped raise awareness of the new clinical model among patients, clinicians, and even health policy makers. They have played an important role in the development of intelligent support for decision-making, clinical practice, and public health policies.

According to recent reviews, the features of PM research are as follows. First, some reviews reported by top experts in the field of PM research discuss the foundation, techniques, applications, and perspectives of this new discipline [4,20,29]. Second, PM research involves significant interdisciplinary collaboration. Many scientists, such as those specializing in clinical medicine, clinical oncology, systems biology, or biochemistry, are focused on the development of this new field. Advanced technologies, such as next-generation sequencing (NGS), molecular imaging, omics (genomics, proteomics, metabolomics, and microbiomics), nanotechnology, big data, and artificial intelligence, have been applied to laboratory tests in PM to achieve more accurate results [6,7,30-34]. Third, the scope of the PM model has been expanded from clinical oncology to noncancer disciplines. This strategy has led to several innovations in the diagnosis and treatment of mental illness, cardiovascular disease, asthma, and inflammatory bowel disease [35-38]. Finally, the reviewers also emphasized the issues to be solved in the development of PM, such as technological bottlenecks, patient privacy, and ethical challenges [39-41]. The information provided in the reviews made a large contribution to the global acknowledgment of PM.

The Rationale for the Study

Research on PM is still increasing, and some important discoveries have already been beneficial to patients. However, there is still a long way to go in the utilization of PM. How can interdisciplinary researchers start studies? What type of public policy really makes sense regarding the field? How can funders ensure that investment works effectively? All these decisions should be made based on knowledge of PM, so great efforts have been made to describe the nature of this new field. The aim of our study was to address the following problems:

1. What is the distribution of topics in PM research?
2. What is the correlation structure of topics in PM research?
3. What are the evolutionary venations and development trends of PM research?

Methods

Data Collection and Processing

According to previous studies, papers in the Web of Science Core Collection (WOSCC) can represent the status of medical science, including PM; therefore, we chose WOSCC as our data source. Data processing is shown in Figure 1.

Papers were collected from the WOSCC that covered the period from 1999 to 2018. In this study, initial retrieval was conducted using “precision medicine,” “P4 medicine,” “personalized medicine,” and “stratified medicine” as terms in the field of Topic to guarantee a recall ratio. It included the document types of Article, Review, and Proceedings. The retrieval strategy is illustrated as follows: TOPIC: (“precision medicine”) or TOPIC: (“personalized medicine”) or TOPIC: (“individualized medicine’’) or TOPIC: (“P4 medicine’’) or TOPIC: (“stratified medicine”). Refined by: Document Types: (Article or Review or Proceedings Paper) Timespan: 1999 to 2018. Indices: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, and IC.

A total of 25,573 publications were retrieved, and their bibliographic records were downloaded through the function Save to Other File Formats provided by WOS. Next, a text file containing all records in Tab-delimited (Win) format was obtained. In general, records without keywords data were excluded. Meanwhile, publications not containing the search terms above in Title (the TI field) or Keywords (the DE field) were identified as unrelated to PM research [31] and also excluded. Thus, 10,177 records were selected as the final data sample, and there were 17,818 unique keywords. Figure 2 shows the number of papers in the sample by year. The PM research started proliferating in 2000, and the number of related papers increased each year.

In this study, mainstream keywords of high frequency were selected for further analysis, that is, based on their coword network. The largest connected component extracted from the whole coword network represents the mainstream research directions of one field [42]. After several rounds of testing of the largest connected component using social network analysis, keywords with a frequency of ≥20 were selected. The sum of the frequency of these words accounts for 52.19% (24,492/46,921) of the total that can represent mainstream research of PM. Meanwhile, keywords were normalized to ensure consistent treatment of the singular and plural forms of words, unifying the synonyms and clarifying the homonyms. For example, the items “target, targets, targeting, target-Specific” were replaced by “Targeted Therapy.” Keywords with a frequency of less than 20 were merged into broader terms. For example, “Hematopoietic Stem Cell” (frequency 6) is replaced by “Stem Cell” (frequency 107). General items which were too broad to be of practical connotation, such as “medicine,” “research,” and “mechanism,” were removed. With the replacement, 244 related words with a frequency greater than 20, which were collected as the basic sample for analysis, were used in this study.

Figure 1. Search procedure for documents in precision medicine research. DE: descriptor; TI: title; WOSCC: Web of Science Core Collection.

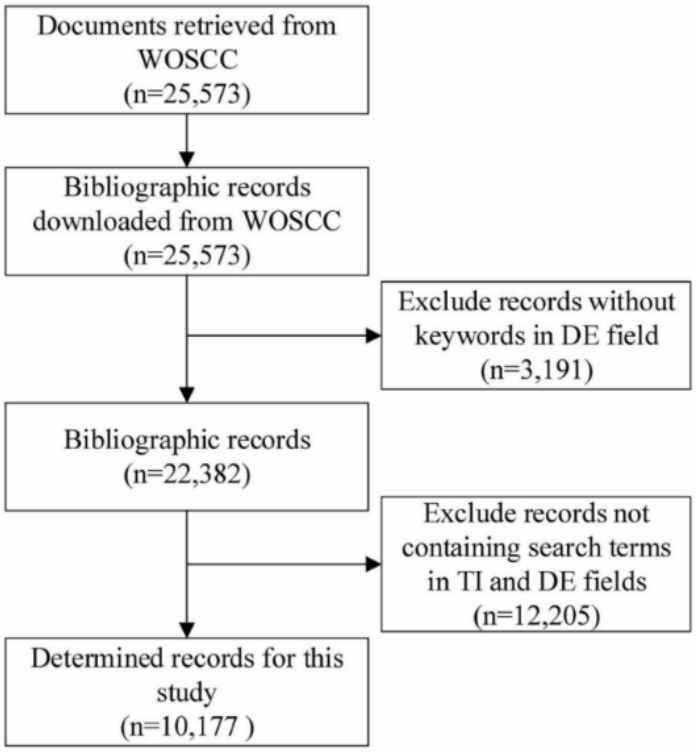
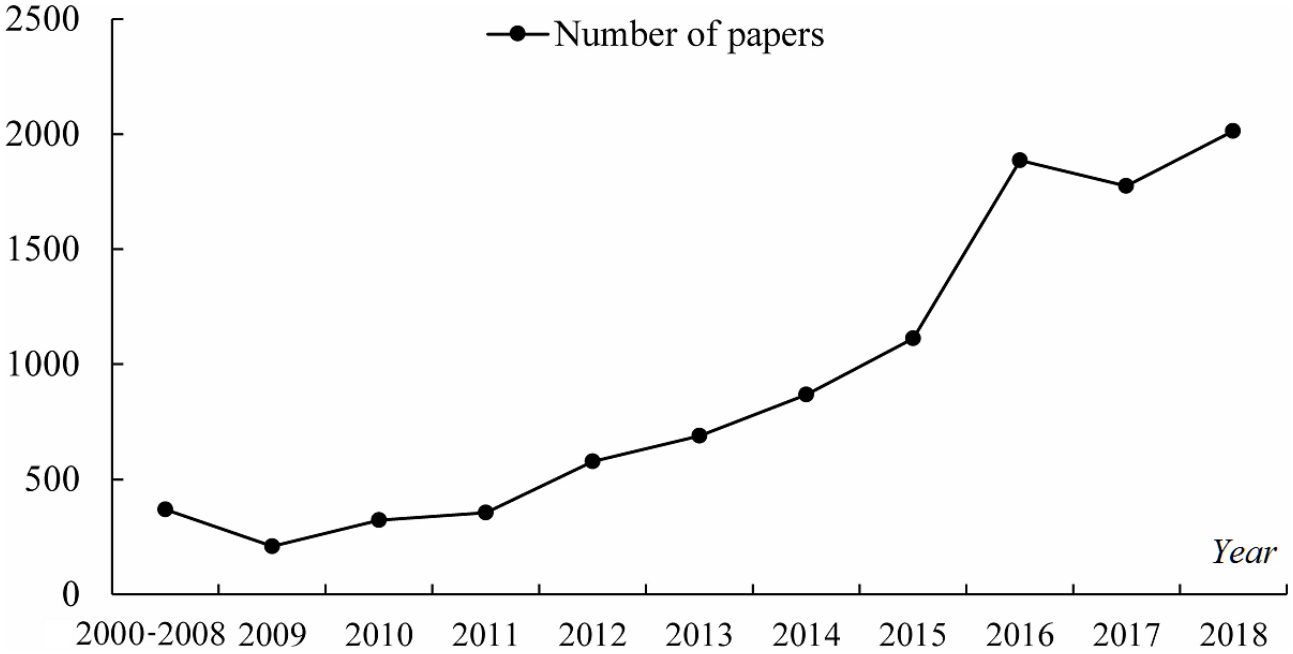


Figure 2. Basic statistics of the sample papers from 2000 to 2018.



Methodology and Tools

Keywords in a paper provide an adequate description of its contents. A study on the correlation of keywords can reveal connotations of contents [43]. Co-occurrence [13] of words (coword) is a kind of correlation in connotation, that is to say, two keywords co-occurring within the same document (eg, paper) are an implication of correlation between topics which they refer to [44]. The high frequency of co-occurrence of each keyword pair means a high degree of correlation between them. Coword analysis has been proven to be effective in identifying main themes and revealing the intellectual structure and patterns of a research field in many previous studies [42,43]. In this study, we used coword analysis that combines both social network analysis tools and scientific mapping tools to analyze the research field of PM. These tools were used to detect and visualize its overall research structure and patterns, conceptual subdomains (thematic communities), and thematic evolution.

Social Network Analysis

Many methods have been used to conduct coword analysis, including the method of social network analysis. It is derived

from mathematical graph theory, which computes indicators of a coword network and identifies characteristics of the whole network and an individual network [45]. SCI2, version 1.2 beta (Cyberinfrastructure for Network Science Center, Indiana University, Bloomington, Indiana, United States), is an effective bibliometric tool to extract items (eg, keywords, authors, and institutions) from bibliography records of articles or other structured research literature [46]. Its feasibility and effectiveness has been widely proven in previous studies [31]. In this study, the bibliographic record file was imported into SCI2 to obtain statistical data of keywords and coword network data. Coword network data include both keywords and their links with the respective weights. The weight of a keyword is its frequency of occurrence and that of the link between the keyword pair is its frequency of co-occurrence.

As unconnected or uncorrelated keywords cannot reflect main thematic subdomains and as what we focused on is the largest component [47], we used SCI2 to exclude isolated nodes and extract the largest component of the coword network [48,49]. Network indicators of the largest component of the coword network were then calculated using Pajek [50], including centralization (centrality), density, and the clustering coefficient. Network indicators of the whole network or nodes can be used to identify the overall intellectual structure and patterns of one research field as well as a keyword's characteristics, such as power, stratification, ranking, and inequality, in the network [31].

Centralization measures the overall characteristics of global network, degree centralization measures the centripetal degree, and closeness centralization measures the proximity degree between any 2 nodes in the network. Its high level equals the close distance between any 2 nodes on the whole. Betweenness centralization indicates the degree of correlation between any 2 nodes through a third one (bridge), and its high level equal the high possibility of correlation through a bridge. Similarly, centrality, the individual network indicator, measures the capacity of one node in network. High degree centrality of one node indicates that it is central in the network and is correlated to many other nodes. It also indicates its powerful capacity of influence and control. High closeness centrality equals the capacity of one node that correlates others as short as possible or directly correlates others. High betweenness centrality equals the powerful role as a bridge to correlate other 2 nodes. Density measures the correlation strength within the network [51]. It means the higher the density, the more mature the research field. The clustering coefficient indicates the possibility that keywords are clustered into a contrasting group [52].

In addition, community detection is an effective method to discover research directions or subfields according to the correlation structure of the network [53]. The Louvain algorithm embedded in Pajek, the most common algorithm used to detect communities, was also used to detect communities in this study [54]. Different communities, including highly correlated keywords, represent different research directions or subfields.

Visualization and Evolution Analysis

Visualization is an important method to intuitively display the intellectual structure of coword correlation, the thematic evolution of a research field, and even the comparative development trends of subfields [55,56]. After repeated comparison of several visualization tools, VOSviewer, version 1.6.13, Centre for Science and Technology Studies, Leiden University, Leiden, Netherlands), was found to enable better visualization of topological networks and was selected to conduct the visualization in this study, including the overall network with communities and the individual networks [57]. The research themes of PM have been evolving over time. We divided bibliographic records chronologically and imported them into Cortext [58]. Evolutionary trends of the keyword community were visualized, allowing for a layout of the dynamics as depicted by tubes in an alluvial model [59].

A strategic diagram indicates the comparative status and evolutionary trends of subfields of one research field. It is a two-dimensional (2D) map in which the x-axis represents centrality and the y-axis represents density [60]. The origin of the axes is determined by the average centrality and density. Centrality can be understood as a measurement of importance and the degree of core in the whole research network. Density can be understood as a measurement of maturity of a theme's development. A total of 4 quadrants in a strategic diagram represent different meanings. Themes in Quadrant 1 are central and developed, with both high centrality and high density; in Quadrant 2, they are highly developed but isolated, with high density and low centrality; in Quadrant 3, they are marginal and isolated (emerging or declining), with both low centrality and low density; and in Quadrant 4, they are central with a trend toward high centrality but low density. Therefore, the developing status and trends of themes or research communities can be predicted by a strategic diagram.

Results

Themes Involved in Precision Medicine Research

In this study, a total of 17,818 keywords were extracted from the sample, and the total frequency was 47,883. The frequency distribution conforms to the power law distribution with an exponent of –1.32 (Figure 3). This shows that the frequency of very few keywords is very high, whereas most keywords are of extremely low frequency. The results indicate that the subject trends in current PM research are obvious, and researchers are inclined to focus on a few major themes and pay less attention to most other themes in the PM field.

Table 1 lists the 100 most frequent keywords, the sum of the frequencies of which accounts for up to 39% of the total frequency. The keywords are so typical and representative in research topics that they can be considered as the mainstream themes of PM research in the past decade. It is interesting that the proportion of the 10 most frequent keywords is 14.2%. Biomarkers and Genomics are the first echelon; Cancer, Therapy, and Genetics are the second echelon; Drug, Target Therapy, Pharmacogenomics, Pharmacogenetics, and Molecular belong to the third echelon. The findings highlight the core and mainstream of PM research topics and show the imbalanced status of PM research as well.

Figure 3. The distribution of the keyword frequency in PM research.

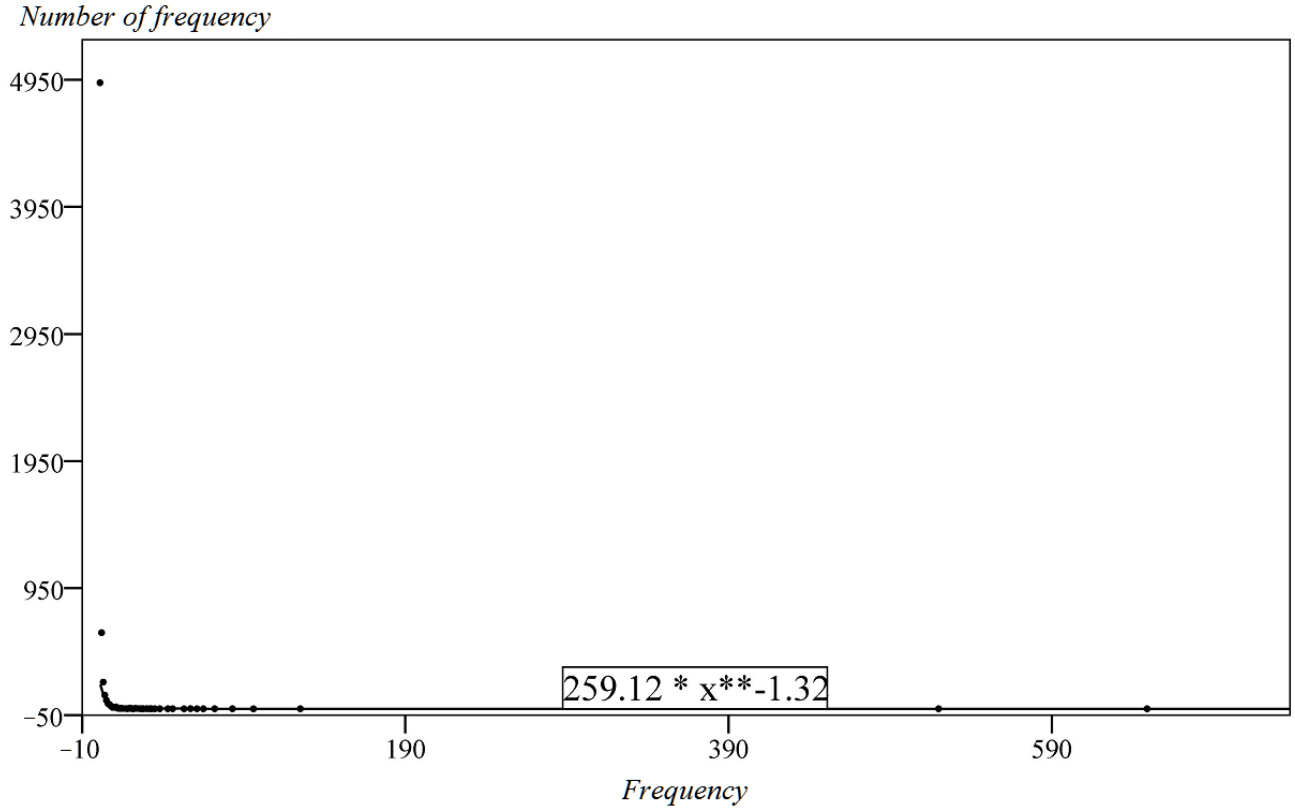


Table 1. Top 100 keywords in precision medicine research.

Number	Keywords	Frequency
1	Biomarkers	1018
2	Genomics	970
3	Cancer	851
4	Therapy	731
5	Genetics	684
6	Drug	549
7	Target Therapy	510
8	Pharmacogenomics	508
9	Pharmacogenetics	475
10	Molecular	357
11	Breast Cancer	333
12	NGS[a]	314
13	Tumor	296
14	Prediction	287
15	Mutation	281
16	Clinical Trials	268
17	Gene	259
18	Sequencing	242
19	Imaging	239
20	Diagnostics	223
21	Proteomics	211
22	Prognosis	209
23	DNA	195
24	Phenotype	187
25	Oncology	185
26	SNP[b]	173
27	Omics	170
28	Pharmacology	165
29	Metabolism	160
30	Lung Cancer	160
31	Bioinformatics	151
32	Asthma	148
33	Chemotherapy	147
34	Immunotherapy	146
35	Stem Cell	143
36	MicroRNA	137
37	Epigenetics	137
38	Prostate Cancer	135
39	Genetic Test	132
40	EGFR[c]	129
41	Risk	129

Number	Keywords	Frequency
42	Inflammation	126
43	GWAS[d]	125
44	Polymorphism	124
45	Colon Cancer	123
46	Immune	123
47	Nanotechnology	122
48	PET[e]	122
49	Translation Medicine	120
50	NSCLC[f]	119
51	Heterogeneity	118
52	Big Data	118
53	Systems Biology	117
54	Machine Learning	117
55	Protein	115
56	Pathology	115
57	Genotype	114
58	Ethics	111
59	Health Care	110
60	Drug Development	110
61	Pharmacokinetics	109
62	Drug Delivery	108
63	RNA	105
64	Diagnosis	105
65	Prevention	103
66	Biobank	103
67	Biology	102
68	Patients	100
69	Diabetes	100
70	Theranostics	100
71	Metabolomics	97
72	Liquid Biopsy	97
73	Screening	97
74	Depression	96
75	Classification	95
76	MRI	93
77	Molecular Imaging	92
78	Brain	91
79	Decision Support	91
80	Electronic Health Records	89
81	Systems Medicine	89
82	Resistance	88
83	Cardiology	87

Number	Keywords	Frequency
84	Clinical Medicine	87
85	Circulating Tumor Cell	86
86	Pancreatic Cancer	84
87	Companion Diagnostics	83
88	Nanoparticle	83
89	Toxicity	81
90	Radiology	81
91	Mass Spectrometry	79
92	Drug Resistance	77
93	Clinical Practice	75
94	Microarray	74
95	Cell	73
96	Metastasis	72
97	Molecular Diagnostics	70
98	Education	70
99	Gastric Cancer	70
100	Gene Expression	70

[a]NGS: next-generation sequencing.

[b]SNP: Single Nucleotide Polymorphisms.

[c]EGFR: epidermal growth factor receptor.

[d]GWAS: genome-wide association studies.

[e]PET: positron emission tomography.

[f]NSCLC: non–small cell lung cancer.

Correlation Network Analysis of Precision Medicine Research

Network Indicators of the Correlation Structure of the Themes

The 244 keywords (frequency above 20) in the study generate a total of 9178 edges, which constitute a keyword correlation network. It is known that the network is the largest connected component, indicating that a relatively consistent mainstream direction has been formed in PM studies in recent years. As shown in Table 2, the degree centralization and closeness centralization of the keywords are relatively high, indicating that the overall network is more concentric, and most of the keywords are clustered, centering on a few core keywords. We also discovered that the keywords in the network tend to be directly correlated rather than indirectly correlated. The path between keywords is short and tends to be directly correlated with the core words. Therefore, it can be concluded that a few core words have very strong control of the entire network. According to the characteristics listed above, we can draw the following conclusions: current PM research is very centralized, the difference between the core words and noncore words is obvious, and the main themes are formed around the core words. However, the lower betweenness centrality also indicates that most of the keyword correlations in PM research can form direct correlations without other words working as bridges. Combined with higher clustering coefficients, it can be observed that there are multiple thematic directions in this PM study with a large degree of difference. The correlation between keywords within the subject direction is higher than that between the other directions. Finally, the overall network is closely correlated, equaling a high network density. This result means that PM research has formed a systematic, relatively mature research pattern.

In the same way, the indices used to describe each keyword (degree centrality, closeness centrality, and betweenness centrality) represent their position and role in the network. As shown in Table 3, Table 4, and Table 5, the keywords, such as Biomarkers, Genomics, Therapy, Cancer, Genetics, Drug, Prediction, Pharmacogenomics, Target Therapy, and Molecular, occupied the top 10 positions on the lists of degree centrality and closeness centrality. It is particularly worth mentioning that the orders of the keywords in the lists of degree centrality and closeness centrality are identical, and both indicators are of high value. These words are clustered around a large number of keywords to a large extent, which are themselves directly correlated with other keywords. This indicates that these keywords are very important, are in the core position, and have a strong influence on the entirety of PM research. In contrast, the value of betweenness centrality is low. Instead of using an intermedia or a bridge word, most keywords are directly correlated, and the connection path is short. Interestingly, the ranking of the keyword “Therapy” is significantly improved in

the list of betweenness centrality compared with its position in the lists of degree centrality and closeness centrality. It indicates that "Therapy" plays an important role of bridging other keywords in the overall network in PM research.

Table 2. The statistics of the correlation network in precision medicine research.

Indicators	Value
Number of nodes	244
Number of edges	9178
Average degree	75.2295
Network all degree centralization	0.6214
Network all closeness centralization	0.6685
Network betweenness centralization	0.0277
Network clustering coefficient	0.4843
Density	0.3096

Table 3. Top 10 keywords in terms of degree centrality.

Ranking	Keywords	Degree
1	Biomarkers	225
2	Genomics	222
3	Therapy	220
4	Cancer	215
5	Genetics	213
6	Drug	208
7	Prediction	184
8	Pharmacogenomics	183
9	Target therapy	177
10	Molecular	172

Table 4. Top 10 keywords in terms of closeness centrality.

Ranking	Keywords	Closeness
1	Biomarkers	0.9310
2	Genomics	0.9205
3	Therapy	0.9135
4	Cancer	0.8967
5	Genetics	0.8901
6	Drug	0.8741
7	Prediction	0.8046
8	Pharmacogenomics	0.8020
9	Target therapy	0.7864
10	Molecular	0.7739

Table 5. Top 10 keywords in terms of betweenness centrality.

Ranking	Keywords	Betweenness
1	Therapy	0.0305
2	Biomarkers	0.0304
3	Genomics	0.0304
4	Drug	0.0289
5	Genetics	0.0283
6	Cancer	0.0260
7	Pharmacogenomics	0.0182
8	Prediction	0.0166
9	Target therapy	0.0148
10	Gene	0.0135

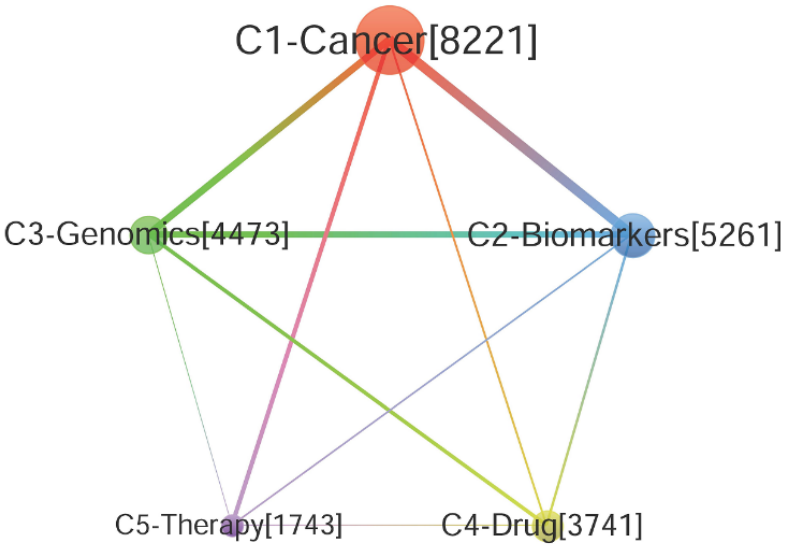
The Themes of the Correlated Communities

On the basis of community detection in the coword network, PM research has focused on 5 theme communities or research subdirections in the last decade. These communities are visualized as Figure 3 in the next section. Modularity (0.2077) [61] of community detection indicates a good result to distinguish topic communities in PM research. Each community has a strong internal correlation, and the distinction between them is obvious. These communities are as follows: C1-Cancer (including Target Therapy, Molecular, Breast Cancer, NGS, Tumor, Mutation, Clinical Trials, Gene, and Prognosis), C2-Biomarkers (including Prediction, Diagnostics, Proteomics, Phenotype, Omics, Metabolism, Bioinformatics, Asthma, and Inflammation), C3-Genomics (including Genetics, Sequencing, Epigenetics, Genetic Test, Risk, Genome-Wide Association Studies, Translation Medicine, Ethics, and Health Care), C4-Drug (including Pharmacogenomics, Pharmacogenetics, Single Nucleotide Polymorphisms, Pharmacology, Polymorphism, Genotype, Drug Development, Pharmacokinetics, and Depression), and C5-Therapy (including Therapy, Imaging, Stem Cell, Nanotechnology, positron emission tomography [PET], Drug Delivery, Theranostics, MRI, Molecular Imaging, and Brain). According to the research scale, PM studies can be divided into 3 levels: Level 1, C1, is the largest level; Level 2, including C2, C3, and C4, is the medium scale; and Level 3, C5, is the smallest. On the basis of the results above, the study of PM mainly focused on Cancer, Biomarkers, Genomics, and Drug in the past decade. More importantly, these themes represent the mainstream direction of PM studies; however, the C5-Therapy community is still weaker than the other 4 communities.

Visualization of the Theme Correlation Network

The structural characteristics of PM research need to be further assessed by the visualization of its coword networks. As shown in Figure 4, each node represents one theme community or research subdirection. The size of the node, determined by the sum of the frequency of all words in the community, represents the scale of this direction. Each edge represents the correlation between the theme communities. Thicker edges indicate greater correlation strengths and a greater influence between communities. In general, C1-Cancer, C2-Biomarkers, C3-Genomics, and C4-Drug have formed a closely related and stable research structure; however, C5-Therapy, loosely correlated with the communities mentioned above, is considered an isolated and marginal research direction. It is noteworthy that the C1-Cancer community has the highest correlation with other communities, highlighting its important position and influence in the entire PM research field. Particularly, C1 has shown that its correlation strength with C2 and C3 is at the highest level. The 3 communities above can be regarded as core directions of PM research, which have the strongest interaction with and influence on each other. In addition, the correlation between the C1 and C5 communities is also strong, indicating interactions between the 2 research directions of Cancer and Therapy, as well as Genomics and Drug.

Figure 4. Correlation structure of theme communities in PM research.

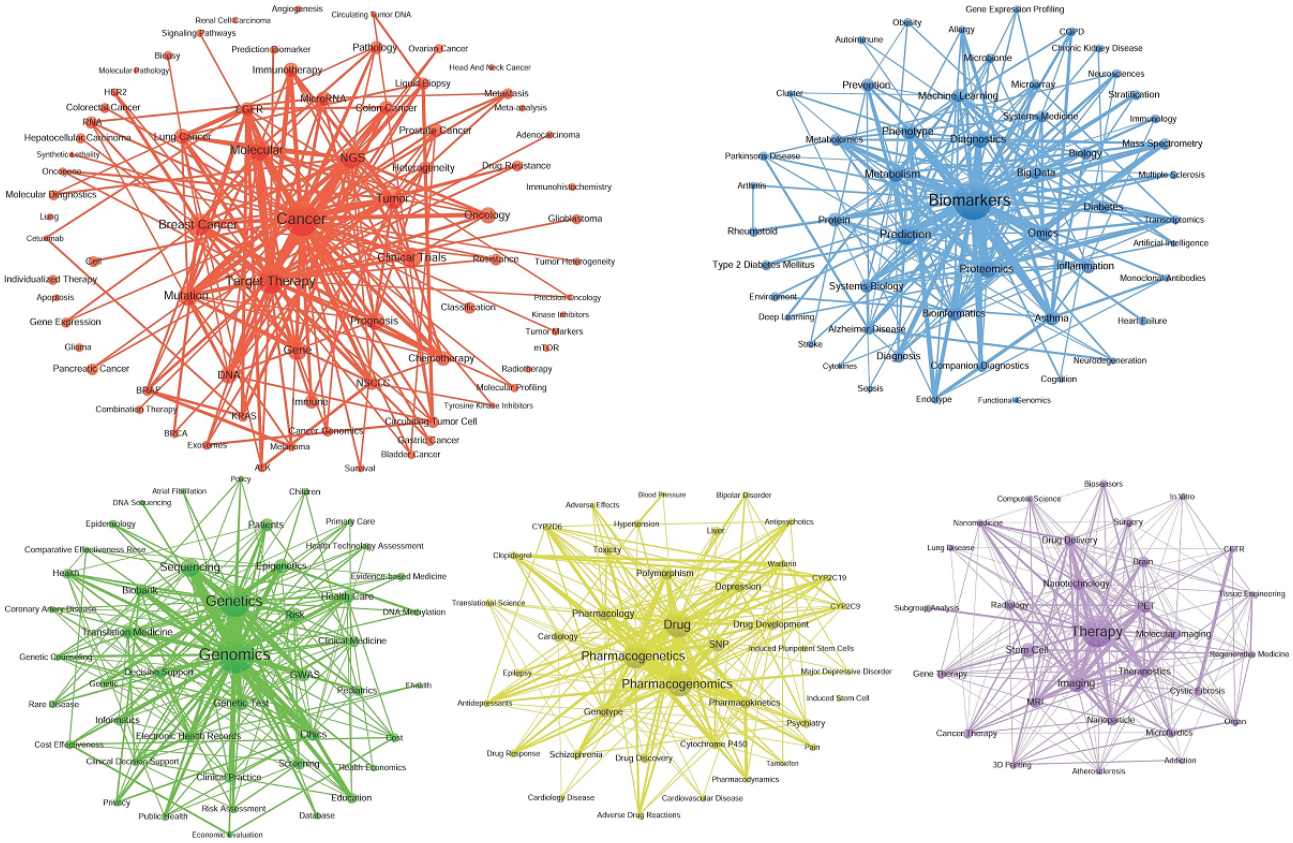


Furthermore, in terms of the internal correlation of the research themes community (Table 6 and Figure 5), especially regarding the indices of the average degree and density, the degree centrality of C1-Cancer and C2-Biomarkers is the highest. Second, C3-Genomics and C4-Drug are subcore research themes, whereas C5-Therapy is a self-contained research theme in PM research but in a marginal position. Finally, C1-Cancer theme community is the most closely correlated within the community and the most mature subject direction in this PM research. The other thematic communities are also closely correlated within them and have a relatively mature development. Overall, the density of all PM research topic communities is higher than the overall density of the coword network. Each research direction has been self-contained and well-developed. However, the strength of the correlation between communities is much weaker than that within the community. The results show that PM research directions are significantly differentiated and that the correlations and interactions between communities are generally insufficient.

Table 6. Indicators of 5 theme communities in precision medicine research.

Community	Number of nodes	Number of edges	Total frequency	Average degree	Density
C1-Cancer	76	1535	8221	82.8026	0.5386
C2-Biomarkers	53	652	5261	78.6792	0.4731
C3-Genomics	45	469	4473	70.7778	0.4737
C4-Drug	40	385	3741	68.375	0.4936
C5-Therapy	30	211	2743	65.7667	0.4851

Figure 5. A total of 5 theme communities in precision medicine research. EGFR: epidermal growth factor receptor; NGS: next-generation sequencing; NSCLC: non–small cell lung cancer.



Evolution of and Trends in Precision Medicine Research

The bibliographic data were divided with the year as the unit of time, and an evolution graph was generated to reveal the evolutionary patterns of PM research. In addition, based on centrality and density, theme communities were graphed in a strategic graph (a 2D map). The relative status and development trend of each theme community in the PM research were revealed.

Evolution Venation of Precision Medicine Research

Overview

To clearly show the development, the evolution of PM research was divided into 2 stages, namely, Stage 1 (2009-2013) and Stage 2 (2014-2018), as shown in Figures 6 and 7. Tubes are colored in each year to represent different topic communities. They are linked because of overlap in keywords in 2 adjacent years, and the evolution venations will be generated with the same color as shown in the figures. According to variations in the topics, such as overlapping, differentiation and fusion of topics, and isolation, we aimed to determine the developing trends of PM. In the past ten years, the continuity in PM themes has generally been good, and a consensus on research directions has formed. Moreover, research in PM has deepened and expanded, especially in Stage 2 (2014-2018), where PM research has maintained good continuity. In the same period, there was more differentiation and integration of the research areas; the interaction between the subjects of the studies was also more pronounced.

Figure 6. The evolution of theme communities of precision medicine research over time (2009-2013). ALK: ALK Receptor Tyrosine Kinase; BRAF: v-raf murine sarcoma viral oncogene homolog B1; BRCA: BReast CAncer gene; EGFR: Epidermal growth factor receptor; HER2: Receptor tyrosine-protein kinase erbB-2; NGS: Next-generation sequencing; KRAS: Ki-ras2 Kirsten rat sarcoma viral oncogene homolog; mTOR: The mammalian target of rapamycin; NSCLC: Non–small cell lung cancer.

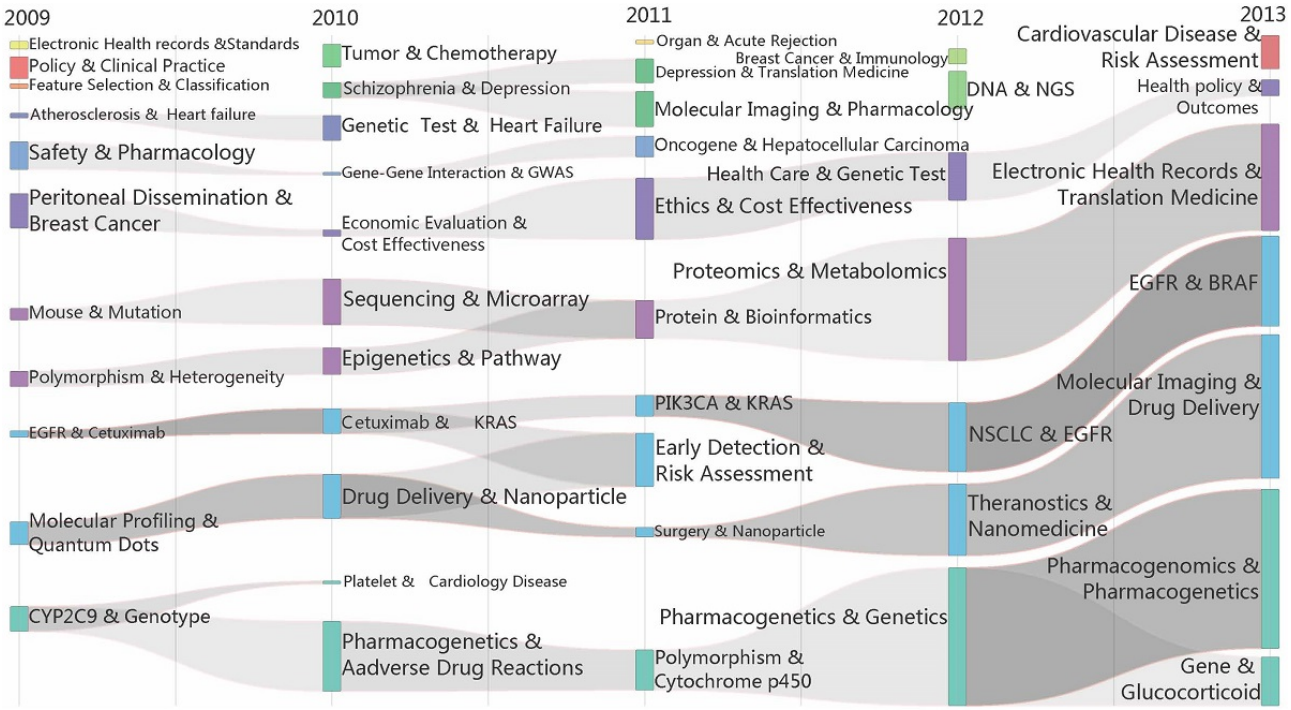
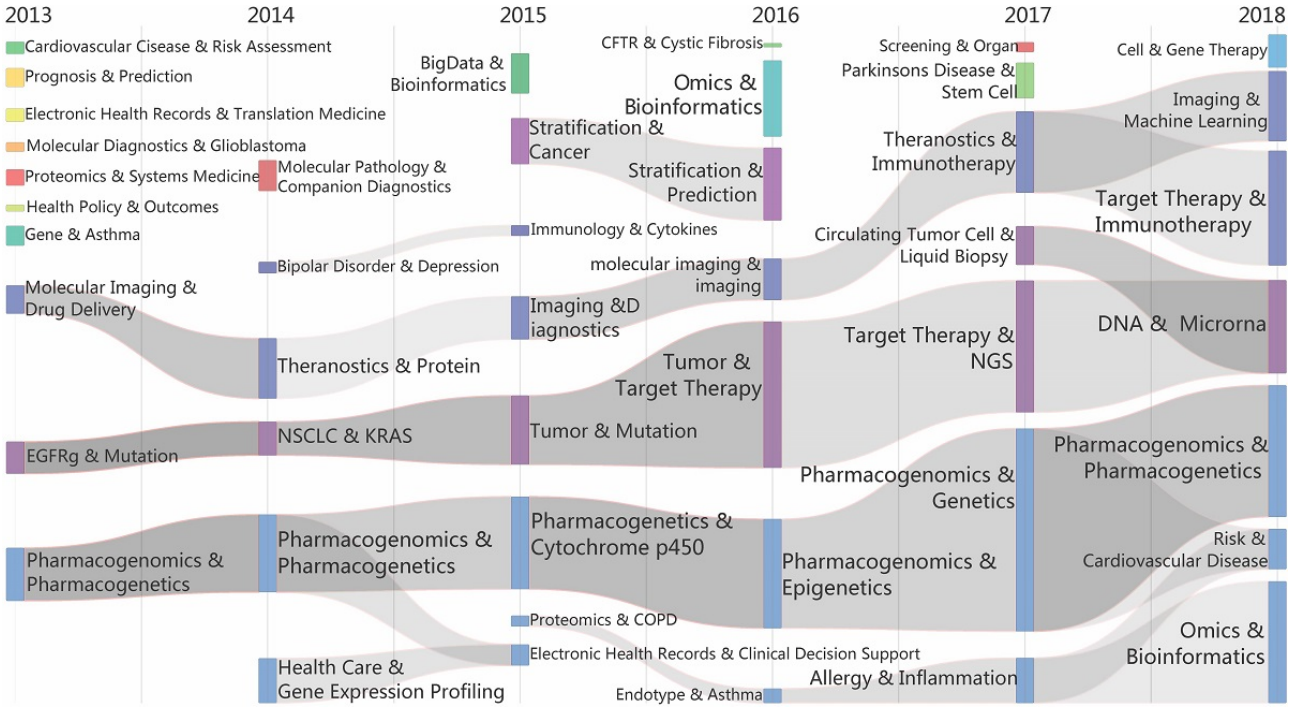


Figure 7. The evolution of theme communities of precision medicine research over time (2013-2018). ALK: ALK Receptor Tyrosine Kinase; BRAF: v-raf murine sarcoma viral oncogene homolog B1; CYP2C9: Cytochrome P450 2C9; EGFR: Epidermal growth factor receptor; GWAS: genome-wide association study; KRAS: Ki-ras2 Kirsten rat sarcoma viral oncogene homolog; NGS: Next-generation sequencing; NSCLC: Non–small cell lung cancer; PIK3CA: phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha.



Stage 1 (2009-2013)

First, there are 4 obvious thematic evolutions: the Pharmacogenomics and Pharmacogenetics venation (including Pharmacogenomics, Genetics, Polymorphism, Adverse Drug Reactions, and CYP2C9), the epidermal growth factor receptor (EGFR) and v-raf murine sarcoma viral oncogene homolog B1 (BRAF) venation (including Molecular Imaging, Drug Delivery, non–small-cell lung cancer [NSCLC], and Ki-ras2 Kirsten rat sarcoma viral oncogene homolog [KRAS]), the Proteomics and Metabolomics venation (including Sequencing, Bioinformatics, and Translation Medicine), and the Ethics and

Cost-Effectiveness venation (including Health Care, Genetic Test, Health Policy, and Breast Cancer).

Each venation is independent and less differentiated, and the internal system for the theme communities is relatively mature. The Pharmacogenomics and Pharmacogenetics venation and the EGFR and BRAF venation are larger scale, so they can thus be considered the 2 important research directions in this period. The evolution of some themes, such as Schizophrenia and Oncogenes, has been interrupted, which may be due to the lack of continuous concern about such subjects or their integration into other subjects. We also find that there are a few isolated themes during different periods, such as Policy, Clinical Practice, Tumor, Chemotherapy, Organ, and NGS. Owing to strong internal correlation, these themes have been clustered as a research direction. However, such studies have not yet formed a systematic and continuous direction.

Stage 2 (2014-2018)

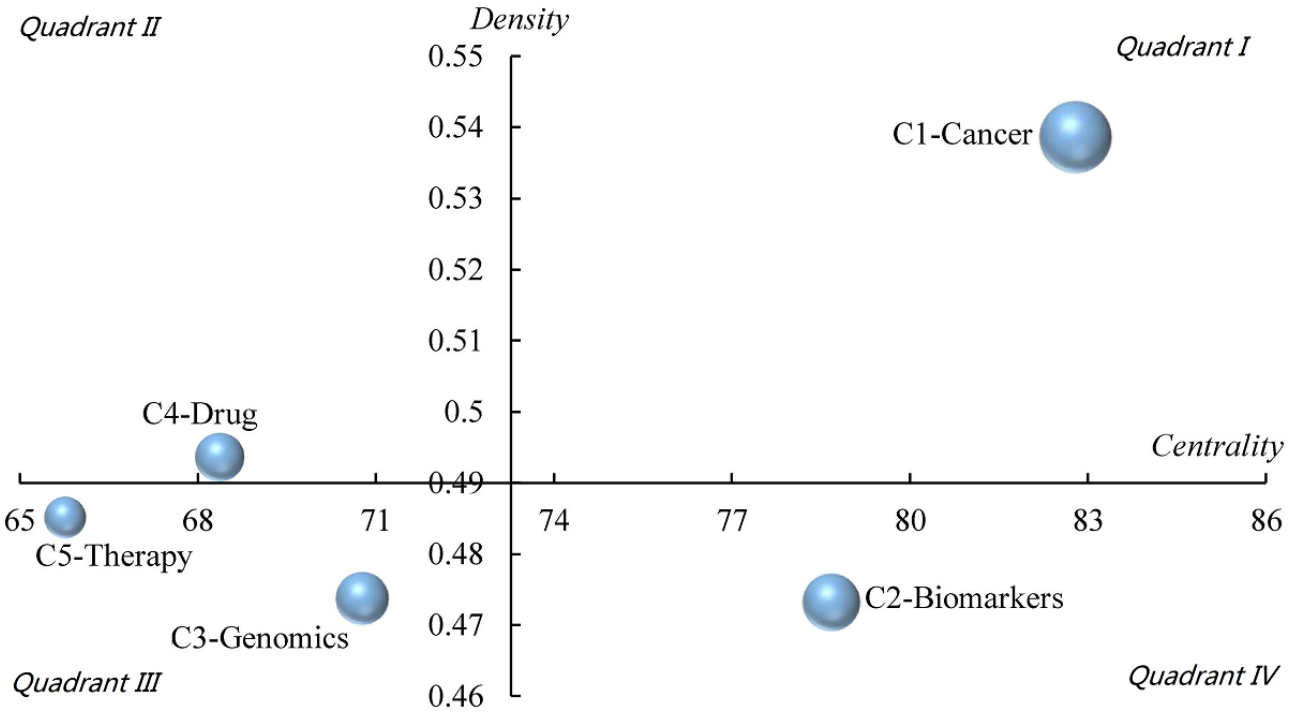
We performed an independent analysis for the years 2013 and 2018 to discover the continuity between 2013 and 2014. There are many overlapping thematic communities in these 2 years as well as overlapping research themes, such as EGFR and BRAF, Molecular Imaging and Drugs, and Pharmacogenomics and Pharmacogenetics, which exhibit good continuity. Overall, the sustainability and stability of PM research in this stage are better than that in Stage 1. Research on PM in terms of themes is more concentrated, which indicates the more consistent and mature direction of progression.

According to the evolutionary graph, there are 3 major research themes at this stage: Molecular Imaging and Drug Delivery, EGFR and Mutation, and Pharmacogenomics and Pharmacogenetics. First, the Molecular Imaging and Drug Delivery venation includes Theranostics, Diagnostics, Immunotherapy, and Machine Learning. The EGFR and Mutation venation includes NSCLC, KRAS, Tumor, Target Therapy, NGS, DNA, and MicroRNA. The Pharmacogenomics and Pharmacogenetics venation includes Cytochrome P450, Epigenetics, Cardiovascular Disease, Omics, and Bioinformatics. Simultaneously, Stratification and Prediction and related topics have also formed an independent evolutionary venation. Although small in scale, they have also become a self-contained system. However, there are also discontinuous evolutions and isolated topics at this stage, such as the evolution of Bipolar Disorder, which was interrupted in 2015. In this period, Parkinson Disease, Stem Cell, and Big Data finally become isolated research themes rather than evolutionary venations.

Development Trends in Precision Medicine Research

The theme community in the PM study is distributed in the strategic map according to centrality and density (Figure 8). On the basis of indicator analysis of the theme communities, we found that C1 is in the first quadrant. Its centrality degree and density are relatively high, indicating that it is the core and mature direction of PM research. C4 is in the second quadrant, with low centrality degree and high density. It can be considered to be the mature direction, but not the core of PM research. In the third quadrant, C3 and C5, with both low centrality and density, are not the core directions and are not mature. However, C3 is close to the origin of density, which means it has the potential to be the core of PM research and that it will develop into a mature community. C2 is in the fourth quadrant, the centrality of which is high, but the density is low. C2 can be considered the core direction, but it is generally immature or involves too many topics.

Figure 8. The relative development status and trends of 5 theme communities in the strategic diagram.



Discussion

Principal Findings

Based on the results, it is possible for us to better understand the main research directions of PM research and accurately evaluate its importance, maturity, and interactions. First, we determined that overall work in PM research is unbalanced but that the theme community is balanced. As PM was newly born as an independent academic subject, researchers paid most of their attention to only a few popular words, such as Biomarkers, Genomics, Cancer, Therapy, Genetics, Drug, Target Therapy, Pharmacogenomics, Pharmacogenetics, and Molecular. The words mentioned above can be classified into the following categories: The applied subject (Cancer), The associated technology and research (Biomarkers, Genomics, and Genetics), pharmacology (Pharmacogenomics), and clinical practice (Treatment, Risk Prediction, Molecular Target Treatment, and Diagnosis). These words not only reflect areas of scientific concern, but more importantly, they indicate the major research directions of PM. However, we also found that the attention paid to most research themes is relatively dispersed. We could speculate that the current status of PM research is possibly as follows: (1) the most mature application of PM is in the subject of Oncology; (2) scientists are interested in discovering Biomarkers, mainly using genomics and genetic methods; (3) pharmacology is an important interdisciplinary field involved with PM, with the aim to make drug utility safer and more efficient; and (4) PM is widely used in Clinical Medicine, including for consulting, diagnosis, and treatment (especially molecular target treatment).

With the visualization of the coword network, we found that the themes were more inclined to be clustered around other popular minority keywords. Thus, the theme communities, both well-layered and balanced-scaled, were finally formed. The communities included C1-Cancer, C2-Biomarkers, C3-Genomics, C4-Drug, and C5-Therapy. According to the analysis of correlation between the theme's communities, we can draw the following inferences: C1-Cancer, as the largest community, indicates that the application of PM in Clinical Oncology might already be mature. The other directions, such as technical studies and Clinical Medicine, are widely associated with Cancer. C2-Biomarkers is the second largest group and plays a key role as the basis of PM research. Scientists still strive for biomarker discovery with various techniques and for the transformation of these discoveries into clinical therapeutics and the prediction of clinical outcomes [61,62]. Owing to significant progress in Genomics and Pharmacology, C4-Drug community, as an independent community, indicates special concern by both pharmacologists and clinicians. In this area, scientists are trying to explore the genetic correlation of Pharmacology and Genomics. These findings will be the foundation of PM, improving drug efficacy and safety [63,64]. It is also noteworthy that the development time of C4-Drug is short, but the fastest. Significant progress has been made in Genomics, Pharmacological Dynamics, Pharmacology, and Metabolomics, and these disciplines are playing an increasingly important role in the field. Although C3-Genomics is relatively isolated and not at the core of PM research, Genomics is one of the most important methods of detecting Biomarkers and is still widely used in various fields of PM [65]. Its decline is due to the application of new technologies, such as high-throughput Omics [66] and Molecular Imaging technology [67]. C5-Therapy, independent but of the smallest scale, indicates that individualized treatment is the ultimate goal of PM, resulting in this aspect gaining the attention of scientists. However, the strategy for treatment is still far from well-developed, which proves the limited scale of the community. According to the major themes included in the C5-Therapy community, individualized treatment mainly involves traditional strategies such as Surgery, Chemotherapy, and Radiology. Interestingly, new methods such as gene therapy, stem cell, and tissue engineering have been available in PM treatment. On the other hand, PET and Molecular Imaging are new technologies that can be applied for stratifications. Through the strategic diagram, C5-Therapy is noncore in PM research; however, we can infer that while the community has not yet matured, it is of great potential.

Through the analysis of the evolution of theme communities over time, PM research has a clear evolutionary and developmental trend. In 2 stages of evolution, we have discovered a large number of well-concentrated evolutionary pathways, which indicates the maturity of PM. The theme community in PM research is well-structured and contains the core and promising directions, such as Biomarkers, Pharmacogenomics, MicroRNA, Imaging, and even Machine Learning. We also identified a dramatic development in techniques and pharmacology directions. It is worth noting that the trend toward PM in nononcology diseases has the potential to become mature, and NSCLC could develop to become an independent and mature venation. It indicates that the application of PM in NSCLC is relatively mature. Clinicians have applied strategies or technologies involved with PM, such as Biomarker, Molecular Imaging, and Pharmacogenomics, to achieve precise treatment [68-70].

Limitations

Our study reveals the structure and developmental trends of PM research from the perspective of keywords and their relationships. To some extent, this study provides insight into PM research; however, there are still limitations to this work. Regarding the research sample, this study used the literature to reveal the development status of PM. This research method could be regarded as a reasonable and cost-effective strategy rather than a comprehensive and accurate way to evaluate the true status of PM research.

Conclusions and Future Directions

Our study reveals the hotspots, structures, evolutions, and developmental trends of PM research in the past 10 years by means of social network analysis and visualization. We also made the following valuable discoveries: (1) using a graph, the network can describe, in detail, the development of PM research; and (2) the network uncovers the relationship between the themes and the intrinsic mechanism about how they interact, which could provide insights into future research directions.

In the future, we will perform data mining on the content of PM-related literature (eg, reports and illness records) to better reveal the condition of the entire network from various perspectives. In terms of research methods, based on previous work, the efficacy of coword analysis has been identified. Our study also validates this research method, and using it, we were able to obtain some valuable discoveries. In future studies, we aim to perform a further, comprehensive assessment of PM research through various perspectives, such as interdisciplinary research and institutes.

Acknowledgments

This study was supported by the National Natural Science Foundation of China Funded Project (71874125), the Ministry of Education in China's Project of Humanities and Social Sciences (18YJA870004), and the Wuhan University Scientific Research Project (2042014KF0164).

Authors' Contributions

JH and XL conceptualized the study and collected and analyzed the data. They participated in all phases of the review. WD and XX assisted with study conception and design, as well as interpretation of data, drafting of the manuscript, and critical revision. All authors contributed to the writing of the manuscript and approved the final version.

Conflicts of Interest

None declared.

References

1. Mohler J, Najafi B, Fain M, Ramos KS. Precision medicine: a wider definition. J Am Geriatr Soc 2015 Sep;63(9):1971-1972. [doi: 10.1111/jgs.13620] [Medline: 26390003]
2. Geyer FC, Lopez-Garcia MA, Lambros MB, Reis-Filho JS. Genetic characterization of breast cancer and implications for clinical management. J Cell Mol Med 2009 Oct;13(10):4090-4103 [FREE Full text] [doi: 10.1111/j.1582-4934.2009.00906.x] [Medline: 19754664]
3. Luttropp K, Lindholm B, Carrero JJ, Glorieux G, Schepers E, Vanholder R, et al. Genetics/Genomics in chronic kidney disease--towards personalized medicine? Semin Dial 2009;22(4):417-422. [doi: 10.1111/j.1525-139X.2009.00592.x] [Medline: 19708993]
4. Aronson SJ, Rehm HL. Building the foundation for genomics in precision medicine. Nature 2015 Oct 15;526(7573):336-342 [FREE Full text] [doi: 10.1038/nature15816] [Medline: 26469044]
5. Kohler I, Hankemeier T, van der Graaf PH, Knibbe CA, van Hasselt JG. Integrating clinical metabolomics-based biomarker discovery and clinical pharmacology to enable precision medicine. Eur J Pharm Sci 2017 Nov 15;109S:S15-S21 [FREE Full text] [doi: 10.1016/j.ejps.2017.05.018] [Medline: 28502671]
6. Kuntz TM, Gilbert JA. Introducing the microbiome into precision medicine. Trends Pharmacol Sci 2017 Jan;38(1):81-91. [doi: 10.1016/j.tips.2016.10.001] [Medline: 27814885]
7. Duarte TT, Spencer CT. Personalized proteomics: the future of precision medicine. Proteomes 2016;4(4):pii: 29 [FREE Full text] [doi: 10.3390/proteomes4040029] [Medline: 27882306]
8. Schwaederle M, Zhao M, Lee JJ, Eggermont AM, Schilsky RL, Mendelsohn J, et al. Impact of precision medicine in diverse cancers: a meta-analysis of phase II clinical trials. J Clin Oncol 2015 Nov 10;33(32):3817-3825 [FREE Full text] [doi: 10.1200/JCO.2015.61.5997] [Medline: 26304871]
9. Printz C. National Institutes of Health releases new guidelines for stem cell research. Cancer 2009 Sep 15;115(18):4043-4044 [FREE Full text] [doi: 10.1002/cncr.24646] [Medline: 19731340]
10. Hollingsworth SJ. Precision medicine in oncology drug development: a pharma perspective. Drug Discov Today 2015 Dec;20(12):1455-1463. [doi: 10.1016/j.drudis.2015.10.005] [Medline: 26482740]
11. He M, Xia J, Shehab M, Wang X. The development of precision medicine in clinical practice. Clin Transl Med 2015 Dec;4(1):69 [FREE Full text] [doi: 10.1186/s40169-015-0069-y] [Medline: 26302883]
12. He Q. Knowledge discovery through co-word analysis. Libr Trends 1999;48(1):133-159 [FREE Full text]
13. Li F, Li M, Guan P, Ma S, Cui L. Mapping publication trends and identifying hot spots of research on Internet health information seeking behavior: a quantitative and co-word biclustering analysis. J Med Internet Res 2015 Mar 25;17(3):e81 [FREE Full text] [doi: 10.2196/jmir.3326] [Medline: 25830358]
14. Chang X, Zhou X, Luo L, Yang C, Pan H, Zhang S. Hotspots in research on the measurement of medical students' clinical competence from 2012-2016 based on co-word analysis. BMC Med Educ 2017 Sep 12;17(1):162 [FREE Full text] [doi: 10.1186/s12909-017-0999-8] [Medline: 28899380]
15. Leung XY, Sun J, Bai B. Bibliometrics of social media research: a co-citation and co-word analysis. Int J Hosp Manag 2017;66:35-45. [doi: 10.1016/j.ijhm.2017.06.012]

16. Li X, Qiao H, Wang S. Exploring evolution and emerging trends in business model study: a co-citation analysis. Scientometrics 2017;111(2):869-887. [doi: 10.1007/s11192-017-2266-5]
17. Chaker AM, Klimek L. [Individualized, personalized and stratified medicine: a challenge for allergology in ENT?]. HNO 2015 May;63(5):334-342. [doi: 10.1007/s00106-015-0004-y] [Medline: 25940007]
18. Sobradillo P, Pozo F, Agustí A. P4 Medicine: the Future Around the Corner. Arch Bronconeumol 2011 Jan;47(1):35-40. [doi: 10.1016/s1579-2129(11)70006-4]
19. Manolio TA, Green ED. Leading the way to genomic medicine. Am J Med Genet C Semin Med Genet 2014 Mar;166C(1):1-7. [doi: 10.1002/ajmg.c.31384] [Medline: 24619573]
20. Ashley EA. The precision medicine initiative: a new national effort. J Am Med Assoc 2015 Jun 2;313(21):2119-2120. [doi: 10.1001/jama.2015.3595] [Medline: 25928209]
21. Ogino S, Nishihara R, VanderWeele TJ, Wang M, Nishi A, Lochhead P, et al. Review article: the role of molecular pathological epidemiology in the study of neoplastic and non-neoplastic diseases in the era of precision medicine. Epidemiology 2016 Jul;27(4):602-611 [FREE Full text] [doi: 10.1097/EDE.0000000000000471] [Medline: 26928707]
22. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 2001 Mar 15;344(11):783-792. [doi: 10.1056/NEJM200103153441101] [Medline: 11248153]
23. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet 2003 May;72(5):1117-1130 [FREE Full text] [doi: 10.1086/375033] [Medline: 12677558]
24. Mancinelli L, Cronin M, Sadée W. Pharmacogenomics: the promise of personalized medicine. AAPS PharmSci 2000;2(1):E4 [FREE Full text] [doi: 10.1208/ps020104] [Medline: 11741220]
25. Shord SS. The role of clinical pharmacology in oncology dose selection: advances and opportunities in personalized medicine. J Clin Pharmacol 2017 Oct;57(Suppl 10):S99-104. [doi: 10.1002/jcph.922] [Medline: 28921640]
26. Roper N, Stensland KD, Hendricks R, Galsky MD. The landscape of precision cancer medicine clinical trials in the United States. Cancer Treat Rev 2015 May;41(5):385-390. [doi: 10.1016/j.ctrv.2015.02.009] [Medline: 25864024]
27. Fiore RN, Goodman KW. Precision medicine ethics: selected issues and developments in next-generation sequencing, clinical oncology, and ethics. Curr Opin Oncol 2016 Jan;28(1):83-87. [doi: 10.1097/CCO.0000000000000247] [Medline: 26569425]
28. Trosman JR, Weldon CB, Douglas MP, Kurian AW, Kelley RK, Deverka PA, et al. Payer coverage for hereditary cancer panels: barriers, opportunities, and implications for the precision medicine initiative. J Natl Compr Canc Netw 2017 Feb;15(2):219-228 [FREE Full text] [doi: 10.6004/jnccn.2017.0022] [Medline: 28188191]
29. McPadden J, Durant TJ, Bunch DR, Coppi A, Price N, Rodgerson K, et al. Health care and precision medicine research: analysis of a scalable data science platform. J Med Internet Res 2019 Apr 9;21(4):e13043 [FREE Full text] [doi: 10.2196/13043] [Medline: 30964441]
30. Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. J Am Coll Cardiol 2017 May 30;69(21):2657-2664 [FREE Full text] [doi: 10.1016/j.jacc.2017.03.571] [Medline: 28545640]
31. Hu J, Zhang Y. Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization. Scientometrics 2017;112(1):91-109. [doi: 10.1007/s11192-017-2383-1]
32. Chen R, Snyder M. Promise of personalized omics to precision medicine. Wiley Interdiscip Rev Syst Biol Med 2013;5(1):73-82 [FREE Full text] [doi: 10.1002/wsbm.1198] [Medline: 23184638]
33. Lu Z, Minko T. Molecular imaging for precision medicine. Adv Drug Deliv Rev 2017 Apr;113:1-2. [doi: 10.1016/j.addr.2017.08.002] [Medline: 28821310]
34. Davis T. Biomedical, Bio-Nano, Personalized Medicine – It's All Nanomedicine to Us!. Aust J Chem 2012;65(1):3-4. [doi: 10.1071/ch11491]
35. Vieta E. [Personalised medicine applied to mental health: precision psychiatry]. Rev Psiquiatr Salud Ment 2015;8(3):117-118. [doi: 10.1016/j.rpsm.2015.03.003] [Medline: 25959401]
36. Krittanawong C. Future physicians in the era of precision cardiovascular medicine. Circulation 2017 Oct 24;136(17):1572-1574. [doi: 10.1161/CIRCULATIONAHA.117.029676] [Medline: 29061572]
37. Oberle AJ, Mathur P. Precision medicine in asthma: the role of bronchial thermoplasty. Curr Opin Pulm Med 2017 May;23(3):254-260. [doi: 10.1097/MCP.0000000000000372] [Medline: 28319473]
38. Fischer S, Neurath MF. Precision medicine in inflammatory bowel diseases. Clin Pharmacol Ther 2017 Oct;102(4):623-632. [doi: 10.1002/cpt.793] [Medline: 28699158]
39. Korngiebel DM, Thummel KE, Burke W. Implementing precision medicine: the ethical challenges. Trends Pharmacol Sci 2017 Jan;38(1):8-14 [FREE Full text] [doi: 10.1016/j.tips.2016.11.007] [Medline: 27939182]
40. Brothers KB, Rothstein MA. Ethical, legal and social implications of incorporating personalized medicine into healthcare. Per Med 2015;12(1):43-51 [FREE Full text] [doi: 10.2217/pme.14.65] [Medline: 25601880]
41. Duffy DJ. Problems, challenges and promises: perspectives on precision medicine. Brief Bioinform 2016 May;17(3):494-504. [doi: 10.1093/bib/bbv060] [Medline: 26249224]

42. Loncar-Turukalo T, Zdravevski E, Machado da Silva J, Chouvarda I, Trajkovik V. Literature on wearable technology for connected health: scoping review of research trends, advances, and barriers. J Med Internet Res 2019 Sep 5;21(9):e14017 [FREE Full text] [doi: 10.2196/14017] [Medline: 31489843]
43. Wei W, Shi B, Guan X, Ma J, Wang Y, Liu J. Mapping theme trends and knowledge structures for human neural stem cells: a quantitative and co-word biclustering analysis for the 2013-2018 period. Neural Regen Res 2019 Oct;14(10):1823-1832 [FREE Full text] [doi: 10.4103/1673-5374.257535] [Medline: 31169201]
44. Hu J, Zhang Y. Research patterns and trends of Recommendation System in China using co-word analysis. Inf Process Manage 2015 Jul;51(4):329-339. [doi: 10.1016/j.ipm.2015.02.002]
45. Hu C, Hu J, Deng S, Liu Y. A co-word analysis of library and information science in China. Scientometrics 2013;97(2):369-382. [doi: 10.1007/s11192-013-1076-7]
46. Börner K. Plug-and-play macroscopes. Commun ACM 2011 Mar;54(3):60-69. [doi: 10.1145/1897852.1897871]
47. Hu J, Zhang Y. Structure and patterns of cross-national Big Data research collaborations. J Doc 2017;73(6):1119-1136. [doi: 10.1108/jd-12-2016-0146]
48. Leydesdorff L, de Moya-Anegón F, Guerrero-Bote VP. Journal maps, interactive overlays, and the measurement of interdisciplinarity on the basis of Scopus data (1996-2012). J Assn Inf Sci Tech 2015;66(5):1001-1016. [doi: 10.1002/asi.23243]
49. Hu J, Huang R, Wang Y. Geographical visualization of research collaborations of library science in China. Electron Libr 2018;36(3):414-429. [doi: 10.1108/el-12-2016-0266]
50. Doreian P, Lloyd P, Mrvar A. Partitioning large signed two-mode networks: problems and prospects. Soc Netw 2013 May;35(2):178-203. [doi: 10.1016/j.socnet.2012.01.002]
51. Callon M, Courtial JP, Laville F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. Scientometrics 1991 Sep;22(1):155-205. [doi: 10.1007/BF02019280]
52. Albert R, Barabási AL. Statistical mechanics of complex networks. Rev Mod Phys 2002 Jan;74(1):47-97. [doi: 10.1103/revmodphys.74.47]
53. Chen G, Xiao L. Selecting publication keywords for domain analysis in bibliometrics: a comparison of three methods. J Inform 2016 Feb;10(1):212-223. [doi: 10.1016/j.joi.2016.01.006]
54. Blondel VD, Guillaume J, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech 2008 Oct;2008(10):P10008. [doi: 10.1088/1742-5468/2008/10/P10008]
55. Muñoz-Leiva F, Viedma-del-Jesús MI, Sánchez-Fernández J, López-Herrera AG. An application of co-word analysis and bibliometric maps for detecting the most highlighting themes in the consumer behaviour research from a longitudinal perspective. Qual Quant 2012;46(4):1077-1095. [doi: 10.1007/s11135-011-9565-3]
56. Chen Y, Fang S. Mapping the evolving patterns of patent assignees' collaboration networks and identifying the collaboration potential. Scientometrics 2014;101(2):1215-1231. [doi: 10.1007/s11192-014-1304-9]
57. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010 Aug;84(2):523-538 [FREE Full text] [doi: 10.1007/s11192-009-0146-3] [Medline: 20585380]
58. Rosvall M, Bergstrom CT. Mapping change in large networks. PLoS One 2010 Jan 27;5(1):e8694 [FREE Full text] [doi: 10.1371/journal.pone.0008694] [Medline: 20111700]
59. Leydesdorff L, Goldstone RL. Interdisciplinarity at the journal and specialty level: The changing knowledge bases of the journal. J Assoc Inf Sci Tech 2014;65(1):164-177. [doi: 10.1002/asi.22953]
60. Leydesdorff L, Park HW, Wagner C. International coauthorship relations in the Social Sciences Citation Index: Is internationalization leading the Network? J Assoc Inf Sci Tech 2014;65(10):2111-2126. [doi: 10.1002/asi.23102]
61. Wang E, Cho WC, Wong SC, Liu S. Disease biomarkers for precision medicine: challenges and future opportunities. Genomics Proteomics Bioinformatics 2017 Apr;15(2):57-58 [FREE Full text] [doi: 10.1016/j.gpb.2017.04.001] [Medline: 28392478]
62. Collins DC, Sundar R, Lim JS, Yap TA. Towards precision medicine in the clinic: from biomarker discovery to novel therapeutics. Trends Pharmacol Sci 2017 Jan;38(1):25-40. [doi: 10.1016/j.tips.2016.10.012] [Medline: 27871777]
63. Lauschke VM, Milani L, Ingelman-Sundberg M. Pharmacogenomic biomarkers for improved drug therapy-recent progress and future developments. AAPS J 2017 Nov 27;20(1):4. [doi: 10.1208/s12248-017-0161-x] [Medline: 29181807]
64. Relling MV, Evans WE. Pharmacogenomics in the clinic. Nature 2015 Oct 15;526(7573):343-350 [FREE Full text] [doi: 10.1038/nature15817] [Medline: 26469045]
65. Carrasco-Ramiro F, Peiró-Pastor R, Aguado B. Human genomics projects and precision medicine. Gene Ther 2017 Sep;24(9):551-561. [doi: 10.1038/gt.2017.77] [Medline: 28805797]
66. Kim D, Kim Y, Son N, Kang C, Kim A. Recent omics technologies and their emerging applications for personalised medicine. IET Syst Biol 2017 Jun;11(3):87-98. [doi: 10.1049/iet-syb.2016.0016] [Medline: 28518059]
67. Wright CL, Binzel K, Zhang J, Knopp MV. Advanced functional tumor imaging and precision nuclear medicine enabled by digital pet technologies. Contrast Media Mol Imaging 2017;2017:5260305 [FREE Full text] [doi: 10.1155/2017/5260305] [Medline: 29097926]

68. Hofman P. ALK in non-small cell lung cancer (NSCLC) pathobiology, epidemiology, detection from tumor tissue and algorithm diagnosis in a daily practice. Cancers (Basel) 2017 Aug 12;9(8):pii: E107 [FREE Full text] [doi: 10.3390/cancers9080107] [Medline: 28805682]
69. Bahce I, Yaqub M, Smit EF, Lammertsma AA, van Dongen GA, Hendrikse NH. Personalizing NSCLC therapy by characterizing tumors using TKI-PET and immuno-PET. Lung Cancer 2017 May;107:1-13 [FREE Full text] [doi: 10.1016/j.lungcan.2016.05.025] [Medline: 27319335]
70. Yin J, Li X, Zhou H, Liu Z. Pharmacogenomics of platinum-based chemotherapy sensitivity in NSCLC: toward precision medicine. Pharmacogenomics 2016 Aug;17(12):1365-1378. [doi: 10.2217/pgs-2016-0074] [Medline: 27462924]

Abbreviations

2D: two-dimensional
BRAF: v-raf murine sarcoma viral oncogene homolog B1
EGFR: epidermal growth factor receptor
HER2: human epidermal growth factor receptor 2
KRAS: Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
NGS: next-generation sequencing
NSCLC: non–small cell lung cancer
P4 medicine: predictive, preventative, personalized, and participatory medicine
PET: positron emission tomography
PM: precision medicine
WOSCC: Web of Science Core Collection

Edited by G Eysenbach; submitted 13.06.18; peer-reviewed by A Mavragani; comments to author 09.10.18; revised version received 07.10.19; accepted 19.10.19; published 04.02.20.

Please cite as:
Lyu X, Hu J, Dong W, Xu X
Intellectual Structure and Evolutionary Trends of Precision Medicine Research: Coword Analysis
JMIR Med Inform 2020;8(2):e11287
URL: https://medinform.jmir.org/2020/2/e11287
doi:10.2196/11287
PMID:32014844

Original Paper

Analyzing Medical Research Results Based on Synthetic Data and Their Relation to Real Data Results: Systematic Comparison From Five Observational Studies

Anat Reiner Benaim[1], PhD; Ronit Almog[1,2], MPH, MD; Yuri Gorelik[3], MD; Irit Hochberg[4,5], MD, PhD; Laila Nassar[6], PharmD; Tanya Mashiach[1], MA; Mogher Khamaisi[3,4,7], MD, PhD; Yael Lurie[5,6], MD; Zaher S Azzam[8,9], MD, FESC; Johad Khoury[8], MD; Daniel Kurnik[5,10], MD; Rafael Beyar[5,11], MD, DSC, MPH

[1]Clinical Epidemiology Unit, Rambam Health Care Campus, Haifa, Israel
[2]School of Public Health, University of Haifa, Haifa, Israel
[3]Department of Internal Medicine D, Rambam Health Care Campus, Haifa, Israel
[4]Institute of Endocrinology, Diabetes and Metabolism, Rambam Health Care Campus, Haifa, Israel
[5]The Ruth & Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel
[6]Clinical Pharmacology and Toxicology Section, Rambam Health Care Campus, Haifa, Israel
[7]Diabetes Stem Cell Laboratory, Rambam Health Care Campus, Haifa, Israel
[8]Department of Internal Medicine B, Rambam Health Care Campus, Haifa, Israel
[9]The Ruth & Bruce Rappaport Faculty of Medicine and Rappaport Research Institute, Technion-Israel Institute of Technology, Haifa, Israel
[10]Clinical Pharmacology Unit, Rambam Health Care Campus, Haifa, Israel
[11]Rambam Health Care Campus, Haifa, Israel

Corresponding Author:
Anat Reiner Benaim, PhD
Clinical Epidemiology Unit
Rambam Health Care Campus
POB 9602
Haifa, 3109601
Israel
Phone: 972 4 7772280
Email: a_reiner@rambam.health.gov.il

Abstract

Background: Privacy restrictions limit access to protected patient-derived health information for research purposes. Consequently, data anonymization is required to allow researchers data access for initial analysis before granting institutional review board approval. A system installed and activated at our institution enables synthetic data generation that mimics data from real electronic medical records, wherein only fictitious patients are listed.

Objective: This paper aimed to validate the results obtained when analyzing synthetic structured data for medical research. A comprehensive validation process concerning meaningful clinical questions and various types of data was conducted to assess the accuracy and precision of statistical estimates derived from synthetic patient data.

Methods: A cross-hospital project was conducted to validate results obtained from synthetic data produced for five contemporary studies on various topics. For each study, results derived from synthetic data were compared with those based on real data. In addition, repeatedly generated synthetic datasets were used to estimate the bias and stability of results obtained from synthetic data.

Results: This study demonstrated that results derived from synthetic data were predictive of results from real data. When the number of patients was large relative to the number of variables used, highly accurate and strongly consistent results were observed between synthetic and real data. For studies based on smaller populations that accounted for confounders and modifiers by multivariate models, predictions were of moderate accuracy, yet clear trends were correctly observed.

Conclusions: The use of synthetic structured data provides a close estimate to real data results and is thus a powerful tool in shaping research hypotheses and accessing estimated analyses, without risking patient privacy. Synthetic data enable broad access

to data (eg, for out-of-organization researchers), and rapid, safe, and repeatable analysis of data in hospitals or other health organizations where patient privacy is a primary value.

(JMIR Med Inform 2020;8(2):e16492) doi:10.2196/16492

KEYWORDS

synthetic data; electronic medical records; MDClone; validation study; big data analysis

Introduction

Background

Access to large databases of electronic medical records (EMRs) for research purposes is limited by privacy restriction, security laws and regulations, and organizational guidelines imposed because of the assumed value of the data. It, therefore, requires approval of the local institutional review board (IRB), but this regulatory process is often time consuming, thereby delaying research and imposing difficulties on data sharing and collaborations. In addition, researchers could apply for a research grant if preliminary data could be extracted and analyzed before making an IRB application, but this is impossible if the data are inaccessible.

Consequently, data anonymization, namely, making reidentification of patients impossible, is required to balance the risk of privacy intrusions with research accessibility. Establishing effective anonymization techniques will promote the future release of data for global access, envisioning democratization of data for all researchers, and facilitate the use of real-world data as a base for study.

One approach for preventing identification of personal records is data masking, namely, removal of identifying information from a dataset, so that individual data cannot be linked with specific individuals. Other techniques include pseudoanonymization, in which a coded reference is attached to a record instead of identifying information, and aggregation, in which data are displayed as totals [1,2]. However, nonaggregation techniques still pose the risk of exposing individuals, as shown by multiple reports [3-7]. In addition to preventing any initial exploration of the data before the IRB approval, once the approval is granted, the researcher may still be blocked by regulatory and ethical barriers for sharing, transferring, and securing the stored data. An alternative approach is the generation of realistic synthetic records comprising the same statistical characteristics and time-dependent properties as the original data such that their analysis yields the same results without mapping the data elements to actual individuals. Thus, synthetic data that are prepared properly can achieve full and irreversible anonymization.

However, creating synthetic data that not only ensure privacy but also retain the information needed for analysis is far from trivial. Kartoun [8,9] proposes a methodology for generating virtual patient repositories, termed electronic medical records bots (EMRBots), based on configurations of population-level and patient-level characteristics. He explains that although such repositories are of high value for training and education, developing computational methods, and assisting hackathons, they cannot serve for studying and predicting real patient outcomes, as their creation does not account for combinations of associations and time-dependent interactions. To reliably mimic EMRs, linear and nonlinear relationships between the variables as well as the temporal arrangement of medical events must be considered.

Other systems for generating synthetic data assume that the data are selected from common distributions that do not comply with the characteristics of real-world medical data and, therefore, may not retain the correlations between multiple variables. Furthermore, they may use prior knowledge of the anticipated relationships, thereby limiting the possibility of true discovery [10-15]. The Synthea system (MITRE Corporation, Massachusetts) [3,14] models care processes and outcomes for several clinical conditions along with their progression. It relies on publicly available datasets and health statistics and synthesizes data according to clinical guidelines and expertise, thereby potentially reflecting ideal scenarios that are not sufficient to replace real EMRs. The Observational Medical Dataset Simulator (OSIM) [15] offers to synthesize data related to diseases and drugs, based on probability distributions estimated from real data, while accounting for time, gender, and age. Relationships are restricted by OSIM to behave in a specific format as reflected by the estimated transitional probability matrix, thereby limiting the ability to reflect other and more complex relationships.

Recently, autoencoders, a technique based on unsupervised deep learning models, has been proposed for synthesizing patient data. By assuming a large enough patient population, autoencoders can learn a representation of the data and then generate a representation that is close to the original input. For instance, medGAN [16] uses real patient records as input to generate high-dimensional discrete samples through a combination of autoencoders and generative adversarial networks. Although this method shows promise in terms of imitating distributional measures and predictions [16,17], it can synthesize only count and binary variables and ignores the longitudinal nature of medical events. Furthermore, a limited privacy risk was observed, and thus, autoencoders cannot yet be considered safe.

This paper studied the validity of synthetic data generated by the MDClone system (Beer-Sheba, Israel), which synthesizes data based directly on the actual real data of interest. The real data is automatically queried from the EMR data lake just before the synthesis. The system was implemented in a number of studies at our institution, Rambam Health Care Campus, located in Haifa, Israel. Our institution is a 1000-bed tertiary academic hospital in Northern Israel and has been using a proprietary EMR system since 2000 (Prometheus, developed by the hospital's department of information technology). Validating

the use of synthetic data for research necessitates a comparison of the results derived from synthetic data with those based on the original data. Previous validation studies on synthetic health data are scarce, of limited scope, and are typically concerned with secondary uses of the data that have minor clinical implications [3,11,14,18,19]. Furthermore, little has been done to compare the statistical results of synthetic data with those of real data [3,14,19]. A more comprehensive validation process concerning meaningful clinical questions and various types of data and outcomes is required for establishing the suitability of synthetic health data for medical research.

We conducted a cross-hospital study to validate the results obtained from synthetic data in various clinical research projects. This paper presents the validation results for five studies conducted at our institution, concerning omission of recommended medication, effect of time to procedure and of hospitalization measures on postdischarge survival, imaging-related risks, and comparison of diabetic treatments. IRB approval to use real data was received, allowing comparative analysis of real vs synthetic data. These studies were used to assess the accuracy and precision of statistical estimates derived from synthetic patient data. The studies represented various population sizes, types of variables and statistical modeling and were based on the hospital's EMR records routinely generated from 2007 to 2017.

The Synthetic Data Generating System

The MDClone system was used in this study for generating synthetic data. This system has been installed in our institute's information technology platform since 2017, and its implementation includes the generation of a structured data lake, a query tool, and a synthetic data generator. The data lake integrates the EMR records with all hospital data sources relating to patient visits, hospitalizations, coded diagnoses, medications, surgical and other procedures, laboratory tests, demographics, and administrative information. The data are presented in an anonymous and standardized format (Health Insurance Portability and Accountability Act of 1996 style). The query engine allows the retrieval of a wide range of variables, in a defined time frame, around an index event. Once an IRB authorization has been granted, the system enables the eligible investigators seamless access to real data structure and analysis with respect to the authorized dataset [20-22]. Otherwise, the system provides the investigator with easy access to synthetic data by the defined query, without revealing the real patient data.

The algorithm used for generating synthetic data is multivariate in nature and generates all variables together, using a covariance measure. It maintains multivariate relationships even on subpopulations of the data (see demonstration in Multimedia Appendices 1-3), as long as they are not too small to expose individual subjects. It does so without assuming any specific form of the underlying distributions and can accept any input, allowing for the discovery of relationships not known before loading the data.

The algorithm treats categorical variables at the first step, ensuring the use of values not unique to a small number of patients. If a subpopulation is identified as unique, such that patients could be identified by certain variables, the values of these variables are censored from the data for these patients. The algorithm then proceeds to extract statistical characteristics from the data, which are used to generate synthetic data with similar properties.

The generation of synthetic data is performed by random sampling from statistical distributions estimated from the original data; thus, each round of data synthesis based on the same query yields a different cohort with similar statistical features. To verify the reliability and validity of the synthetic data, the system produces a report with (1) censoring rate for each variable; (2) a summary of the distribution of each variable, original vs synthetic; and (3) a comparison of all pairwise correlations.

Methods

Validation Methodology

For each participating study, we repeatedly produced five synthetic datasets based on the query to be used to extract the real data. We then statistically analyzed each set and compared the results, namely, the effect point estimates and their uncertainty levels, as reflected by the confidence intervals, with those obtained from the real data. The types of effects compared included proportions, odds ratios, hazard ratios, and survival curves, as obtained by applying the relevant statistical models.

In addition, to evaluate the stability of results obtained from synthetic data, we evaluated the consistency of the estimates across the synthetic sets. Although an initial impression was obtained from observing the results across the five synthetic sets, we repeatedly generated numerous synthetic sets to evaluate the bias and stability of the estimates. To obtain small enough standard errors, 1000 repetitions were used. Bias was defined by the difference between the mean across all synthetic sets and the estimate obtained from the real data. Stability was evaluated by the range of this difference. The bias and stability were evaluated for three of the studies, which represented the types of statistical outcomes addressed in this study, and reflected the common measures used in clinical research: proportions (the Proton Pump Inhibitors [PPIs] Prescription Study), hazard ratios and survival curves (the Percutaneous Coronary Intervention [PCI] and ST-Elevation Myocardial Infarction [STEMI] Study), and odds ratios (the Hypoglycemia Insulin Study).

Generation of Synthetic Data

For each participating study, the following steps were taken throughout the analysis:

- The investigator logged into the system and defined the patient cohort by setting inclusion and exclusion criteria.
- The information required for these patients was defined by a query. An approximation for the number of patients meeting the criteria was then provided by the system. The researcher could define a reference event (eg, the first myocardial infarction event) that could be used to pull data in relative temporal terms (eg, the last hospitalization before the event). Any data included in the hospital's EMR could be requested, provided it was within the access definitions for the researcher, as set by an administrator.

- The cohort with its defined data was extracted and seamlessly converted into synthetic information with the same structure as the original data. A data file was prepared and downloaded, along with a report providing a descriptive comparison between the synthetic data and the original data for each variable.
- The synthetic data were statistically analyzed.
- Following IRB approval, real data were extracted and analyzed using the same analytics.

Participating Studies

A total of five clinical studies conducted in the hospital were selected for the validation process. The studies addressed contemporary topics with important clinical and medical implications. They represented a range of statistical questions, types of analysis, and population sizes that are frequently confronted in hospital research. Tables describing the real populations are provided in Multimedia Appendices 4-7, and synthetic data files are provided in Multimedia Appendices 8 and 9.

The Proportion of Omission of Proton Pump Inhibitor Prescriptions for Gastroprotection

Gastrointestinal bleeding is one of the most common preventable adverse drug events [23,24], and antiplatelet and anticoagulant medications are the most common drugs associated with hospitalization caused by PPI prescription errors [25]. To reduce the risk of gastrointestinal bleeding, guidelines recommend prescribing PPIs to high-risk patients [26,27]. This study assessed the proportion of PPI omission for gastroprotection in patients discharged with prescribed combinations of oral anticoagulants (OACs; warfarin, dabigatran, rivaroxaban, or apixaban) and antiplatelets (aspirin, clopidogrel, prasugrel, or ticagrelor), accounting for additional indications for prophylactic PPI use (age >65 years and concomitant steroid use). In each subgroup, we examined the proportion of patients with recommended administration of concomitant PPIs.

The Effect of Time to Percutaneous Coronary Intervention in ST-Elevation Myocardial Infarction Patients on Death and Heart Failure

This study examined the effect of door-to-balloon time (D2B) among STEMI patients on the occurrence of congestive heart failure (CHF) or mortality, within 180 days of catheterization. According to the guidelines adopted in Israel in 2014, PCI should be performed within 90 min of arrival to the hospital [28]. Kaplan-Meier survival rate estimates were calculated, and the effect of D2B and STEMI-associated factors [29,30] was estimated by a multivariate Cox proportional hazard regression, accounting for other adverse events, such as severe cardiac presentation (cardiogenic shock, cardiac arrest, ventricular fibrillation, ventricular tachycardia, and complete atrioventricular block) and prior ischemic heart disease (IHD; previous coronary artery bypass surgery, myocardial infarction, and PCI). In addition, laboratory test results indicating low hemoglobin (≤10), high creatinine (>1), and high blood urea nitrogen (BUN; >30), as well as potential confounders (age, gender, and year), were all accounted for.

The Impact of Blood Urea Nitrogen on Postdischarge Mortality Among Patients With Acute Decompensated Heart Failure

Acute decompensated heart failure (ADHF) is the leading cause of hospital admission in patients older than 65 years [31]. This study investigated the effect of BUN levels during hospitalization on 3-year mortality after discharge from the hospital among patients with ADHF. Admission and discharge BUN levels were extracted. The predictive value of BUN for mortality was evaluated using multivariate Cox proportional hazard regression, accounting for the number of associated comorbidities. In addition, the levels of brain natriuretic peptide, red cell distribution width, and blood sodium were included in the model as dichotomous variables, in accordance with accepted thresholds.

The Risk of Nephropathy Following Magnetic Resonance Imaging Using Gadolinium-Based Contrast Agents Compared With the Risk Following Computed Tomography–Based Imaging Using Iodine-Based Contrast Agents

Contrast-induced nephropathy (CIN) following iodine-based contrast-enhanced imaging has been widely known as a leading cause of acute kidney injury (AKI) [32-34]. This study aimed to establish the risk of AKI following contrast-enhanced magnetic resonance imaging (MRI) relative to that of contrast-induced computed tomography (CT). We included all adult patients who had undergone a contrast-enhanced CT or MRI. Propensity score matching was used to account for known risk factors for CIN and AKI by applying nearest-neighbor 1:4 matching between MRI and CT patients. Comorbidities such as diabetes and IHD and the target organ of the imaging study were also accounted for. AKI rates were compared by odds ratios calculated from the full data and the matched data by the Fisher exact test and the Mantel-Haenszel test, respectively.

The Risk of Hypoglycemia in Patients With Diabetes Treated by Detemir or Glargine Insulins by Blood Albumin Level

Detemir and glargine are long-acting insulins commonly used for inpatient treatment [35]. However, detemir is albumin bound, raising a concern for increased risk of hypoglycemia for patients with hypoalbuminemia [36,37], and guidelines for treating hyperglycemia do not prefer one insulin over the other [35]. This study assessed the risk of hypoglycemia in patients with low albumin treated with insulin detemir vs glargine. Retrieved data included all adult patients treated with detemir or glargine and laboratory results for albumin, creatinine, and glucose levels during a 5-day time frame. In addition, age, gender, weight, insulin dose, insulin dose-to-weight ratio, home usage of insulin, receiving of short insulin, division of hospital stay, and length of stay were also accounted for. Hypoglycemia risks were estimated by fitting a multivariate logistic regression model that included main effects and second-order interactions as the predictors and hypoglycemia (glucose level <70 mg/dL) as the dependent variable. Variables were selected for the model by a stepwise procedure based on the Akaike Information Criterion.

Results

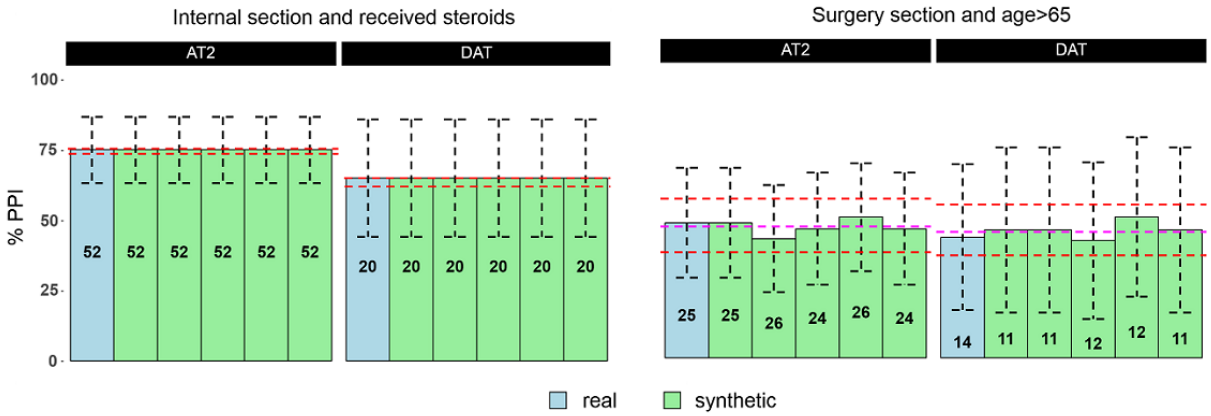
Protein Pump Inhibitors Prescription Study

Between 2007 and 2017, we identified 12,188 patients discharged on OACs, some of whom additionally received a single antiplatelet, either aspirin (n=3953) or P2Y ADP receptor blockers (clopidogrel, prasugrel, or ticagrelor) antiplatelet therapy (n=882), or a double antiplatelet therapy (DAT; n=417).

Comparisons between results obtained from five synthetic sets and the real data are shown in Figure 1. Overall, good predictions of the real data results were obtained from the synthetic data. For most subgroups, such as patients discharged from internal medicine departments with OAC and antiplatelets and receiving concomitant steroids, generating synthetic data did not require censoring, as no observations were found to be unique. Estimates from synthetic data were, therefore, identical to those from real data, regardless of sample size, including their uncertainty levels, as reflected by the confidence intervals (left panel). On the basis of repeated runs of 1000 synthetic sets, the PPI administration proportions for this subgroup were highly stable, as indicated by the nearly zero bias and their very narrow range across the 1000 repeats.

For small subgroups, some instability was observed, as can be readily seen by the estimates obtained from the five synthetic sets (right panel). The estimates' range across the 1000 synthetic sets was wider for those two subgroups (minimum −10.5% and maximum +8.5%). Their overall mean across 1000 sets shows biases of −1.3% and 1.9% for AT2 and DAT, respectively, which are small when compared with the uncertainty level (reflected by the confidence intervals) of the estimates from real data.

Figure 1. PPI administration (%) for patients receiving the clopidogrel, prasugrel or ticagrelor antiplatelet (AT2) or dual antiplatelet (DAT). The total number of patients in the subgroups are given inside the bars. If no censoring was required (left panel – Internal Section patients that received steroids), proportions of PPI administration calculated from the synthetic sets were essentially identical to the proportions in the real data, and their range across 1000 sets (minimum and maximum in red dotted lines) was very narrow. If censoring was required, as in the case of the Surgery Section, results varied across the synthetic sets, and their ranges were wider (right panel – Surgery Section patients older than 65 years). The means across 1000 sets (purple lines) show small biases.



Percutaneous Coronary Intervention and ST-Elevation Myocardial Infarction Study

Between 2013 and 2016, 597 patients diagnosed with STEMI who underwent primary PCI were identified, excluding cases in which more than 6 hours had passed before performing primary PCI or with CHF before intervention. Boolean classifications were used to extract information on patient conditions: the variable *severe cardiac presentation* indicated cardiogenic shock, cardiac arrest, ventricular fibrillation, ventricular tachycardia, or atrioventricular block on admission, and the variable *prior ischemic heart disease* indicated prior coronary artery bypass surgery, myocardial infarction, or PCI.

Survival curves estimated from synthetic data were similar to the curves estimated from real data with little variability between curves obtained from the five synthetic sets (Figure 2) and were within the confidence limits obtained from the real data. The mean curve based on 1000 synthetic sets was similar to the curve obtained from the real data. Hazard ratios for 180 event-free (CHF/death) days are shown in Figure 3. A D2B greater than 90 min revealed no increased risk, based on either the real or the synthetic data. Conclusions were typically consistent between real and synthetic data and across the five synthetic sets. Estimates were also consistent in the uncertainty level (width of confidence intervals). In the case of increased risk with age and borderline significance for a slight increase in risk for patients with prior IHD, as obtained from the real data, some variability was observed. For results with higher confidence, the hazard ratio estimates were more stable. Yet, the bias of the estimate obtained from synthetic data, as estimated by 1000 repeatedly generated synthetic sets, was small when compared with the uncertainty of the estimate from real data. As expected, the stability and the bias of the synthetic results were better for variables with narrower confidence intervals (age group, gender, and year) compared with variables with wider confidence intervals (prior IHD and high BUN).

Importantly, all estimates obtained from synthetic data, for the survival curve and the hazard ratios, were within the 95% confidence limits obtained from the real data, namely, within the range of potential values of the true survival rate and the true hazard ratio.

Figure 2. Kaplan-Meier 180-day event-free (CHF/mortality) survival curves after primary PCI, estimated from the real data with 95% confidence limits (blue) and from five repeatedly generated synthetic datasets (green). Survival curves based on synthetic data were similar to curves based on real data, and the mean curve based on 1000 synthetic sets was similar to the curve obtained from the real data.

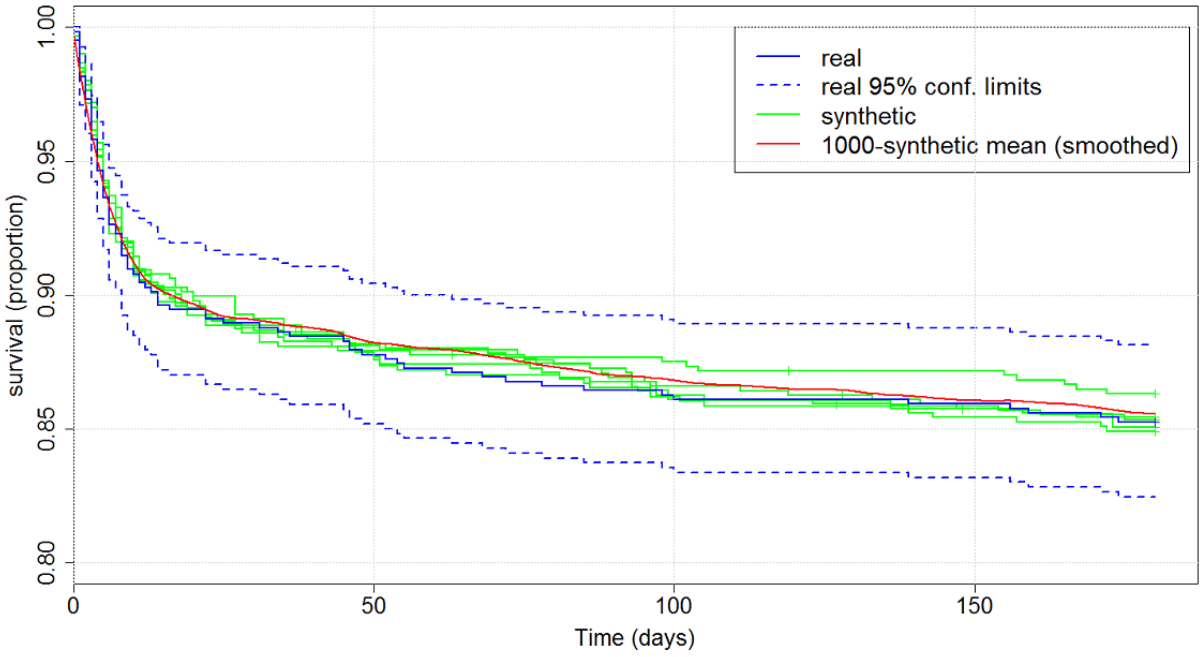
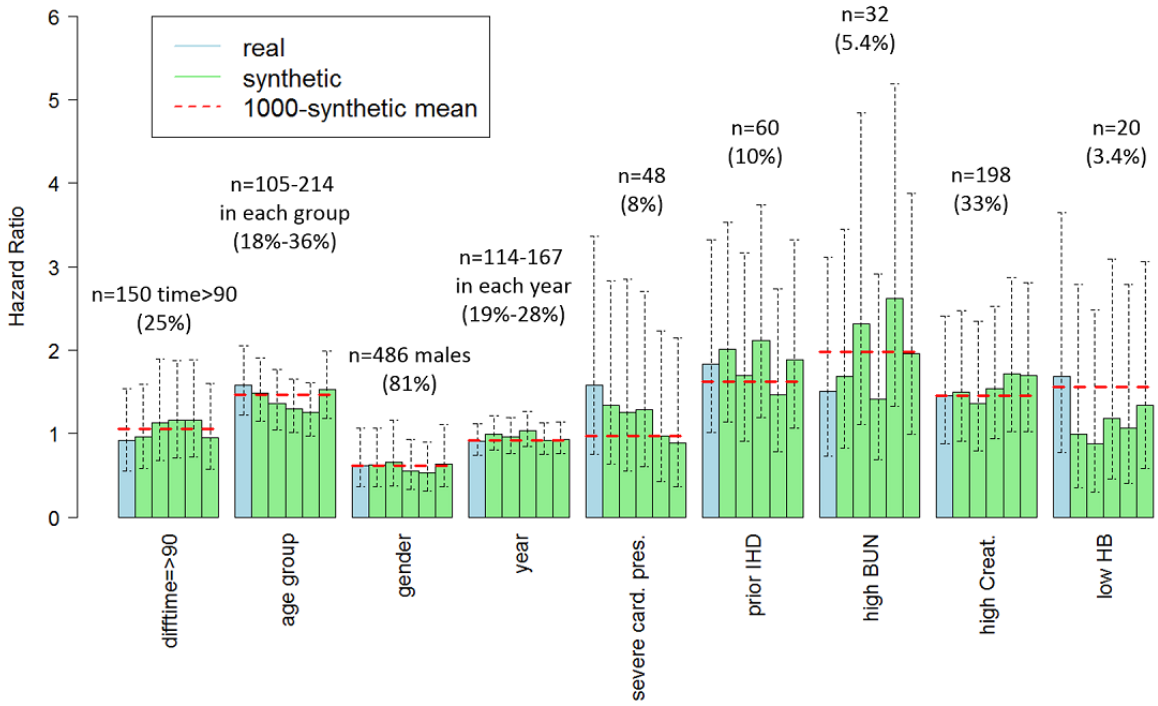


Figure 3. Hazard ratios with 95% confidence intervals for CHF or mortality within 180 days of primary PCI based on real data (blue) and on five synthetic datasets (green). For each variable, the number of cases and percentage in the real data is given. Conclusions were typically consistent between the real and the synthetic data, and across the synthetic sets. In the case of increased risk with age, some variability was observed. The mean result across 1000 synthetic sets (dotted red line) for results with high confidence, was close to the result from the real data, implying small bias.



Blood Urea Nitrogen and Acute Decompensated Heart Failure Study

Between 2007 and 2017, 4590 patients were hospitalized with a primary diagnosis of heart failure and survived to discharge. To limit the number of subgroups, a Boolean classification was used for extracting information on comorbidities instead of specific diagnoses. As shown in Figure 4, Kaplan-Meier 3-year survival obtained from the real data was nearly 60% for an admission BUN of below 30, 44% for BUN of 30 to 39, and 37% for BUN of 40 or above, implying that high admission BUN is a risk marker for mortality within 3 years. Similar estimates were obtained from the five synthetic sets. Hazard ratios relative to the *below 30* group were estimated from the real data as 1.29 for patients with BUN 30 to 39 and 1.67 for patients with BUN 40 or above. Hazard ratios estimated from synthetic data were slightly lower (Figure 5). As in the PCI-STEMI Study, all estimates from synthetic data, for the survival rate and the hazard ratios, were within the confidence limits obtained from the real data.

Figure 4. Kaplan-Meier three-year survival curves by admission BUN level, as estimated from the real data (in blue) and from five repeatedly generated synthetic datasets (in orange). The survival curves estimated from the synthetic sets were very close to the curve estimated from the real data.

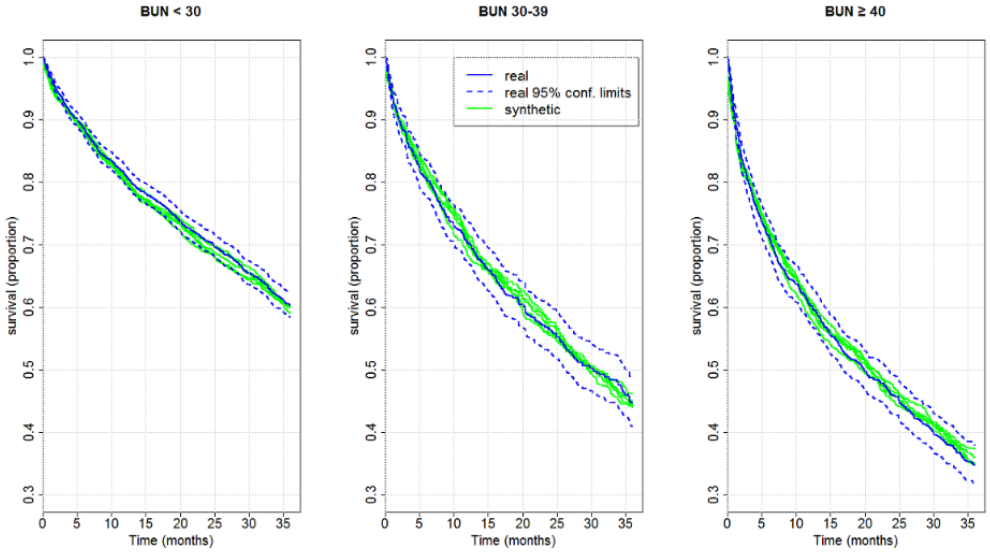
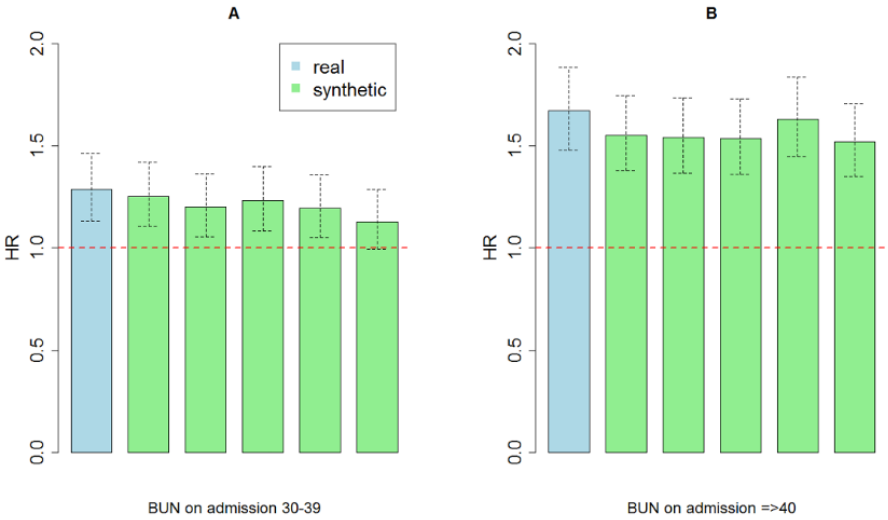


Figure 5. Hazard ratios with confidence intervals by admission BUN level, obtained by Cox proportional hazard regression based on real data and on five synthetic datasets. Hazard ratios relative to the reference group of BUN below 30 based on real data were 1.29 for patients with BUN between 30 and 39 (panel A) and 1.67 for patients with BUN 40 or above (panel B). Hazard ratios estimated from synthetic data were slightly lower. The width of confidence intervals was consistent between the real and the synthetic data, and across the synthetic sets.



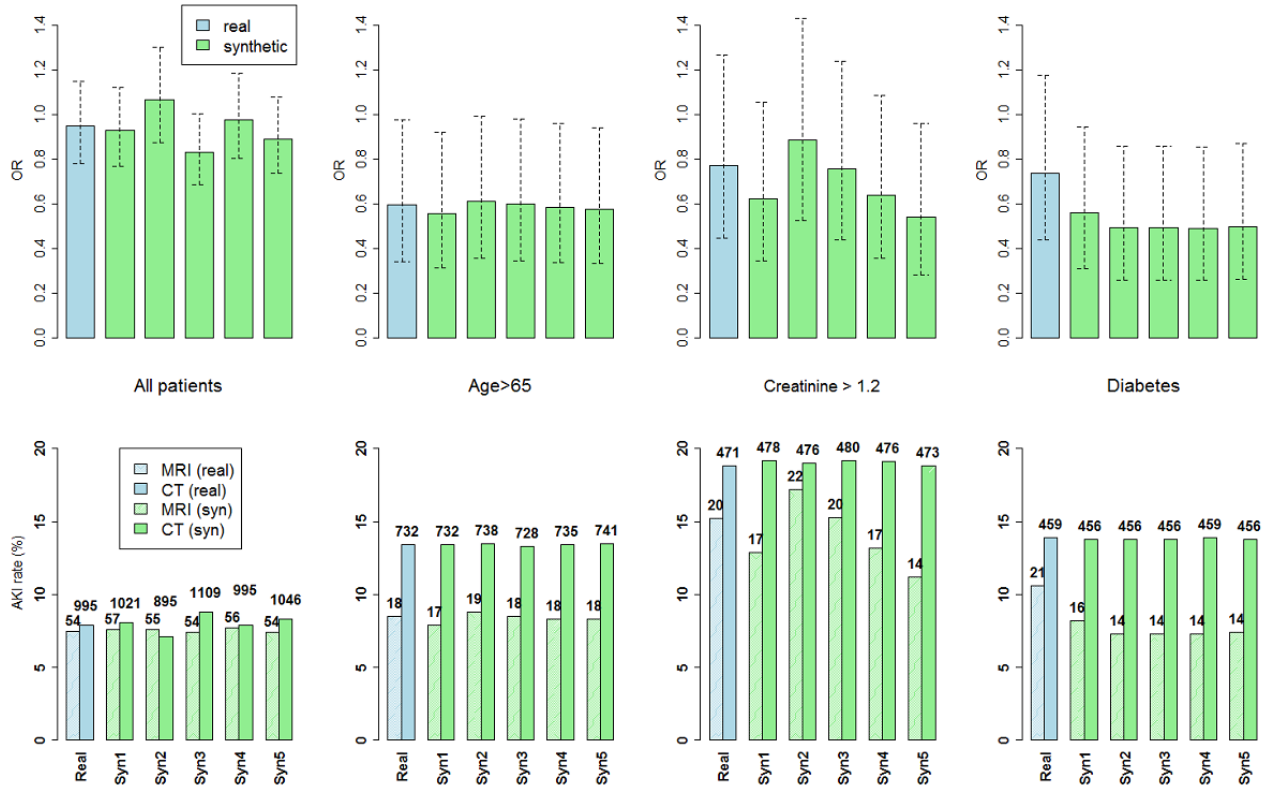
Imaging Nephropathy Study

We identified 718 patients who underwent a contrast-enhanced MRI between 2013 and 2017 and 12,592 patients who underwent CT imaging between 2011 and 2017, excluding patients who underwent additional contrast-enhanced imaging within 3 days around the index imaging. To limit the number of subgroups, diagnoses and drugs were defined as Boolean variables.

Odds ratios obtained from the real data and five synthetic sets are presented in Figure 6. For the relatively large CT group, AKI rates were consistent between the real and the synthetic sets and across the synthetic sets for all patient subgroups. For

the small MRI group, the number of AKI cases per subgroup was only 18 to 21. The AKI rates were well estimated for patients older than 65 years, and the borderline statistical difference remained consistent; the AKI rate estimates were less stable for patients with high creatinine, yet the conclusion of no difference was consistent. For patients with diabetes, AKI rates and odds ratios were lower across all synthetic sets and should, therefore, be interpreted with caution. All odds ratio estimates obtained from synthetic data were within the 95% confidence limits obtained from the real data.

Figure 6. Acute kidney injury (AKI) rates (lower panel) and odds ratios with 95% confidence intervals (upper panel) in four different subgroups for the real data and five repeatedly generated synthetic datasets (Syn1-Syn5). The number of patients in the data for each subgroup is shown above the rate bars. Results obtained from the synthetic data were generally consistent with those obtained from the real data. AKI rates were well estimated for patients older than 65 years of age, and the borderline statistical difference remained consistent; AKI rate estimates were less stable for patients with high creatinine, yet the conclusion of no statistical difference was consistent; Odds ratios for diabetic patients were under-estimated due to under-estimated AKI rates for the very small number of diabetic patients that underwent MRI.



Hypoglycemia Insulin Study

Between 2012 and 2016, 4677 adult patients were hospitalized and treated with detemir (832/4677, 17.78%) or glargine (3844/4677, 82.19%) insulins. The risk curves estimated from the synthetic sets for detemir and glargine treatments across various albumin values (Figure 7) were highly similar to the curves estimated from the real data and consistently indicated the association of detemir use with a higher prevalence of hypoglycemic events in patients with hypoalbuminemia. Figure 8 presents risk predictions for 1000 repeatedly generated synthetic sets, compared with the estimates obtained from the real data. The estimates from all synthetic sets predicted a higher hypoglycemia rate for detemir and were within the confidence limits obtained from the real data. Their bias was −0.003 for detemir and +0.006 for glargine.

Figure 7. Risk predictions with 95% confidence intervals for detemir and glargine insulin treatments for a range of albumin values, based on the real data (top left) and five synthetic datasets (other panels). The risks estimated from the synthetic sets were highly similar to the curves estimated from the real data, and consistently indicated association of detemir use with a higher prevalence of hypoglycemic events in patients with hypoalbuminemia.

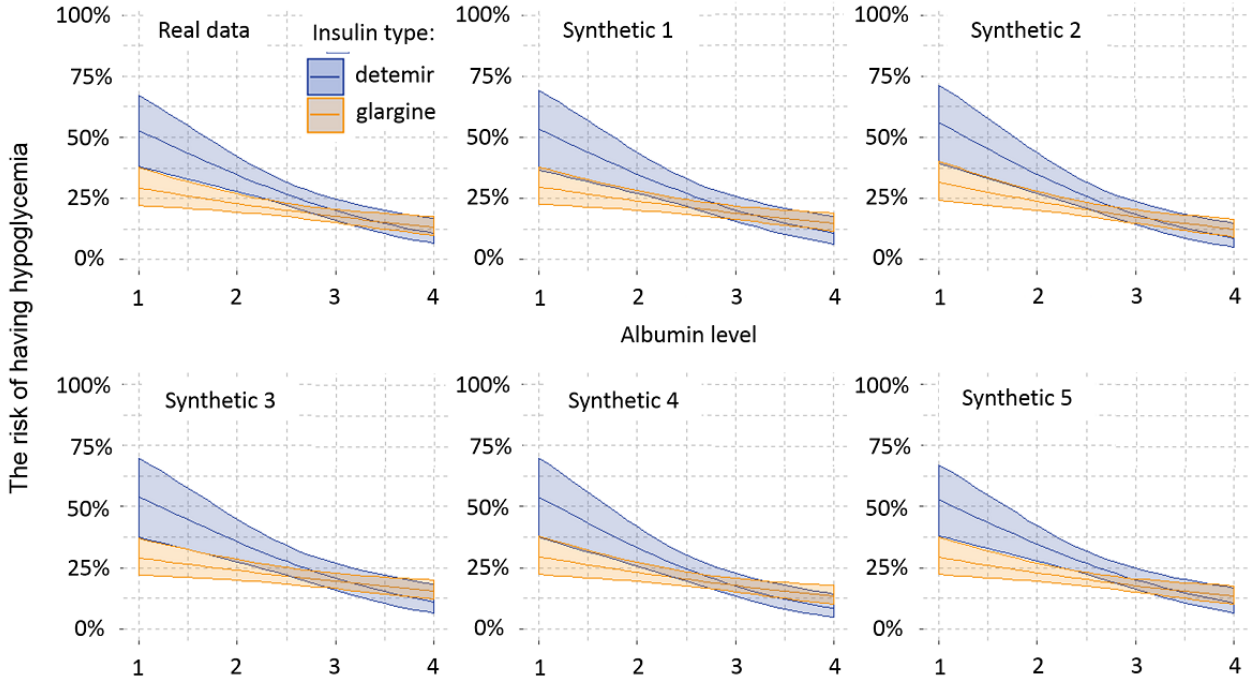
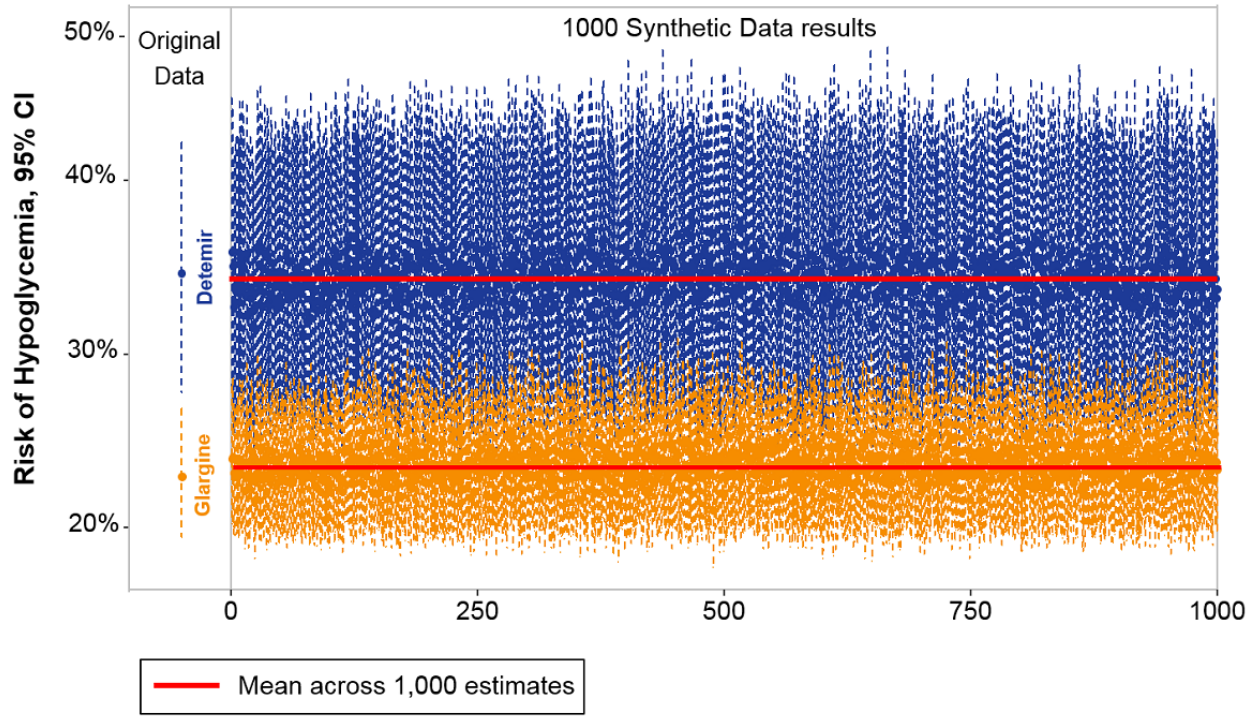


Figure 8. Risk predictions at albumin 2 gr/dL for 1000 repeatedly generated synthetic sets, compared to estimates obtained from the real sets (thin dotted line on the left marks the confidence intervals with the point estimates marked on the line). All synthetic sets predicted a higher hypoglycemia rate for detemir, and all were within the confidence limits of the estimates from the real data. The synthetic data estimates, as showed by their means (thick red lines), are biased from the real data estimates by –0.003 for detemir and by +0.006 for glargine.



Discussion

Principal Findings

The use of synthetic data based on EMR is an approach for obtaining an estimate of real statistical results at a stage when real data are not available for the investigator. This paper examined the validity of statistical results based on synthetic data by comparison with real data for five studies, using medical records from our institution. Our study extended the scope of previous studies and investigated the performance of synthetic data under a variety of medical research questions. We used a system implemented in our institution that transforms the real data to synthetic data, which, when analyzed, provides the investigator with a reasonably accurate estimate of the real data results, and findings based on the synthetic data can be published in accordance with the institution policy. We assumed reliable performance of the system in privacy preservation, yet a future study aimed to investigate and validate issues related to the security and irreversibility of the synthetic data is of high relevance. Furthermore, sharing of synthetic data files that imitate particular real datasets and are generated within the hospital EMR platform must be a strategic decision of the hospital, accounting for concerns that are beyond academic considerations, such as costs of generating the data, timing of its release for sharing, and means of storage and access.

Five clinical studies on different topics, performed by separate research groups, were used for this validation study. The studies varied in population sizes and types of variables and statistical analysis. The validation study showed that the results derived from synthetic data were predictive of real data results. This was demonstrated with high consistency across all clinical studies. When the number of patients was large relative to the complexity and number of variables with very little or no censoring, as in the Hypoglycemia Insulin Study, the system proved itself highly predictive, with strong consistency of results between synthetic and real data, even for analyses involving complex computations and multiple stages such as stepwise logistic regression. Thus, the system can be effectively used to assess results from large data. Furthermore, when no censoring was imposed, precise predictions were obtained for proportions from synthetic data, regardless of sample size, as in the PPI Prescription Study.

For studies based on smaller populations that accounted for confounders and modifiers by multivariate models, such as the PCI-STEMI Study (n=597) and the Imaging Nephropathy Study (n=718), clear trends were still correctly observed by the synthetic data, although the predictions were of moderate accuracy. Nevertheless, these predictions are of high importance for guiding investigators before real data analysis and in generating a predictive hypothesis based on synthetic data that can then be applied to real data.

Several steps should be taken to minimize prediction bias caused by censoring when using synthetic data. Similar to any complex multivariate analysis, researchers should limit the number of variables to the minimum necessary and, when formulating the query, define variables to include information at the minimal required resolution, as in Boolean coding. When adhering to this recommendation, high consistency was achieved in the BUN-ADHF Study, which contained a large number of patients (n=4590) but also many subgroups. In addition, as seen in this study, analysis of multiple synthetic sets can guide the investigator by providing information on the stability of the synthetic results and indicating possible bias.

Comparison With Prior Work

Previous validation studies on synthetic health data primarily considered secondary use of the data, with few medical implications [3,11,14,18,19]. Loong [19] did a limited comparison of statistical results between real and synthetic data, concluding that synthetic data are suitable for exploratory analysis. Walonoski et al [3] compared statistical properties with publicly available statistics for type 2 diabetes and found incorrect results for several variables, such as age at diagnosis, prevalence by racial groups, comorbidity rates, and survival, and acknowledged the need to increase the realistic level of the patient records. In a later paper, Chen et al [14] compared rates obtained from datasets generated by Synthea with publicly reported rates for four health care quality measures, showing inaccuracies that were partly caused by ignoring noncompliance with clinical guidelines and diversity in health care utilization.

Our validation study included a comprehensive validation process concerning meaningful clinical questions and various types of data and outcomes, which represent the scope of studies and type of statistical analysis conducted on hospital records. We used a system that seamlessly synthesizes data based on the actual original data of interest. We compared results obtained from the synthetic data with those obtained from the original data and included analysis of 1000 repeatedly generated synthetic datasets to estimate the bias and stability of the results.

Limitations

Small populations may challenge the synthesis of data by (1) limiting the quality of the estimated statistical characteristics of the original data, particularly for high-dimensional multivariate distributions and outliers, and (2) causing selection bias in the estimates, if censoring of observations is made to prevent patient identification. Yet, as shown in this study, even varied and biased results obtained for very small subgroups, as in the Imaging Nephropathy Study, were still within the confidence limits of the results based on the original data. In addition, although interactions and correlations are preserved by the synthetic data, as shown in this study, high-order and complex relationships can be further investigated for very large study populations that involve hundreds or more variables, where the synthetic data results can also be compared with those generated by autoencoders.

Synthesis of nonstructured data, such as imaging results and free text from medical reports, has not yet been implemented in the synthesis engine and requires structuring of the data using image analysis, natural language processing, or other suitable approaches, enabling the eventual extraction of the statistical characteristics of the data. In addition, for some conditions considered in this paper, such as diabetes and CHF, structured data alone may be incomplete, and thus, extracting information from text can enhance the results on structured data.

Missing values for a particular variable in the original data are treated as a population subcategory by itself, for which statistical characteristics of all other variables are extracted separately. Thus, the synthesized data contain missing values for that subcategory as well. On obtaining the synthetic data, the researcher can decide if and how to impute the missing values, as in the case of real data.

Conclusions

We provide a comprehensive evaluation of the use of synthetic data in comparison with real data, from an EMR data bank of a large academic medical center, based on five clinical studies conducted by five different research groups. In general, results based on synthetic data were highly predictive of those based on real data. Cases and conditions for which prediction may be nonprecise or biased were discussed and typically result from either censoring applied by the system to protect patient anonymity or data samples too small for quality estimation. Synthetic data, interpreted with an understanding of its limitations, are a powerful tool to guide clinical data analysis and research and allow for rapid, safe, and repeated analysis of routine data in a hospital setting and other health organizations where patient privacy is imperative.

Acknowledgments

The authors wish to thank Sara Tzafrir and Idan Sipori from the Rambam Information, Computerization, and Communications Department for contributing to the implementation of the MDClone system, supporting data retrieval, and performing quality assurance. The authors acknowledge the assistance of Deborah Hemstreet, an English editor employed by the Rambam Health Care Campus, in editing this manuscript and preparing it for submission.

Authors' Contributions

AB, advisor for the validation approach and methodology, provided guidance in statistical analysis and results reporting, performed statistical analysis for the PCI-STEMI Study, and wrote the paper. RA was a general advisor for the validation project; medical and methodological advisor for health records retrieval and interpretation; and member of the PCI-STEMI Study group who participated in study design, data collection, manuscript preparation, and review. YG was a member of the Imaging Nephropathy Study group who participated in study design, data collection, data analysis, manuscript preparation, and review. IH was the principal investigator of the Hypoglycemia Insulin Study who participated in study design, data collection, data analysis, manuscript preparation, and review. LN was a member of the PPI Prescription Study group who participated in data collection, data analysis, manuscript preparation, and review. TM performed data collection and statistical analysis for the BUN-ADHF Study. MK was the principal investigator in the Imaging Nephropathy Study who participated in study design, data analysis, manuscript preparation, and review. YL was a member of the PPI Prescription Study group who participated in data collection, data analysis, manuscript preparation, and review. ZA was the principal investigator in the BUN-ADHF Study who participated in study design, data analysis, manuscript preparation, and review. JK was a member of the BUN-ADHF Study group who participated in data collection, data analysis, manuscript preparation, and review. DK was a principal investigator in the PPI Prescription Study who participated in study design, data analysis, manuscript preparation, and review. RB initiated and led the validation project, was the principal investigator in the PCI-STEMI Study, and coedited the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preservation of interactions and associations.
[DOCX File , 546 KB - medinform_v8i2e16492_app1.docx]

Multimedia Appendix 2
Spearman correlation coefficients for all pairs of numeric variables, based on the synthetic data (vertical axis) and the original data (horizontal axis). The correlation is preserved for the wide range of correlations, from negative to positive coefficients.
[PNG File , 170 KB - medinform_v8i2e16492_app2.png]

Multimedia Appendix 3
Boxplot of hemoglobin levels - comparison of MIMIC III (Original) and the synthetic datasets, by patient's age and hematocrit level. The high order correlation between hematocrit level, hemoglobin level and age, is consistent between the original data and the synthetic data. The delicate decline of hemoglobin as age increases, subject to the increase of hemoglobin level with hematocrit level, in general and within age group, is well preserved by the synthetic data.
[PNG File , 115 KB - medinform_v8i2e16492_app3.png]

Multimedia Appendix 4

Data Characteristics Table – PPI Prescription Study.
[DOCX File , 20 KB - medinform_v8i2e16492_app4.docx]

Multimedia Appendix 5
Data Characteristics Table – PCI-STEMI Study.
[DOCX File , 20 KB - medinform_v8i2e16492_app5.docx]

Multimedia Appendix 6
Data Characteristics Table – BUN-ADHF Study.
[DOCX File , 23 KB - medinform_v8i2e16492_app6.docx]

Multimedia Appendix 7
Data Characteristics Table – Hypoglycemia Insulin Study.
[DOCX File , 21 KB - medinform_v8i2e16492_app7.docx]

Multimedia Appendix 8
Synthetic data files and a variable description file - PPI Prescription study.
[ZIP File (Zip Archive), 1995 KB - medinform_v8i2e16492_app8.zip]

Multimedia Appendix 9
Synthetic data files and a variable description file - BUN-ADHF study.
[ZIP File (Zip Archive), 3787 KB - medinform_v8i2e16492_app9.zip]

References

1. Garfinkle SL. National Institute of Standards and Technology. 2015 Oct. De-Identification of Personal Information URL: https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf [accessed 2020-01-20]
2. Graham C. The Information Commissioner's Office (ICO). 2012. Anonymization: Managing Data Protection Risk Code of Practice URL: https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf [accessed 2020-01-20]
3. Walonoski J, Kramer M, Nichols J, Quina A, Moesel C, Hall D, et al. Synthea: an approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. J Am Med Inform Assoc 2017 Aug 30. [doi: 10.1093/jamia/ocx079] [Medline: 29025144]
4. Anderson R. Under threat: patient confidentiality and NHS computing. Drugs Alcohol Today 2006;6(4):13-17. [doi: 10.1108/17459265200600060]
5. Ohm P. Broken promises of privacy: responding to the surprising failure of anonymization. UCLA Law Rev 2010;57:1701 [FREE Full text]
6. El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. PLoS One 2011;6(12):e28071 [FREE Full text] [doi: 10.1371/journal.pone.0028071] [Medline: 22164229]
7. McLachlan S, Dube K, Gallagher T. Using the CareMap with Health Incidents Statistics for Generating the Realistic Synthetic Electronic Healthcare Record. In: Proceedings of the 2016 IEEE International Conference on Healthcare Informatics. 2016 Presented at: ICHI'16; October 4-7, 2016; Chicago, IL. [doi: 10.1109/ichi.2016.83]
8. Kartoun U. arXiv e-Print archive. 2016. A Methodology to Generate Virtual Patient Repositories URL: https://arxiv.org/ftp/arxiv/papers/1608/1608.00570.pdf [accessed 2020-01-20]
9. Kartoun U. Advancing informatics with electronic medical records bots (EMRBots). Softw Impacts 2019;2:100006. [doi: 10.1016/j.simpa.2019.100006]
10. McLachlan S. School of Engineering and Advanced Technology, Massey University. 2017. Realism in Synthetic Data Generation URL: https://mro.massey.ac.nz/bitstream/handle/10179/11569/02_whole.pdf?sequence=2&isAllowed=y [accessed 2020-01-20]
11. Patki N, Wedge R, Veeramachaneni K. The Synthetic Data Vault. In: Proceedings of the 2016 IEEE International Conference on Data Science and Advanced Analytics. 2016 Presented at: DSAA'16; October 17-19, 2016; Montreal, QC. Canada. [doi: 10.1109/dsaa.2016.49]
12. Mwogi TS, Biondich PG, Grannis SJ. An evaluation of two methods for generating synthetic HL7 segments reflecting real-world health information exchange transactions. AMIA Annu Symp Proc 2014;2014:1855-1863 [FREE Full text] [Medline: 25954458]
13. Buczak AL, Babin S, Moniz L. Data-driven approach for creating synthetic electronic medical records. BMC Med Inform Decis Mak 2010 Oct 14;10:59 [FREE Full text] [doi: 10.1186/1472-6947-10-59] [Medline: 20946670]

14. Chen J, Chun D, Patel M, Chiang E, James J. The validity of synthetic clinical data: a validation study of a leading synthetic data generator (Synthea) using clinical quality measures. BMC Med Inform Decis Mak 2019 Mar 14;19(1):44 [FREE Full text] [doi: 10.1186/s12911-019-0793-0] [Medline: 30871520]
15. Murray RE, Ryan PB, Reisinger SJ. Design and validation of a data simulation model for longitudinal healthcare data. AMIA Annu Symp Proc 2011;2011:1176-1185 [FREE Full text] [Medline: 22195178]
16. Choi E, Biswal S, Malin B, Duke J, Stewart WF, Sun J. Generating Multi-Label Discrete Patient Records Using Generative Adversarial Networks. In: Proceedings of Machine Learning for Healthcare 2017. 2017 Presented at: MLHC'17; August 18-19, 2017; Boston, MA, United States URL: http://proceedings.mlr.press/v68/choi17a/choi17a.pdf
17. Zhang Z, Yan C, Mesa DA, Sun J, Malin BA. Ensuring electronic medical record simulation through better training, modeling, and evaluation. J Am Med Inform Assoc 2020 Jan 1;27(1):99-108. [doi: 10.1093/jamia/ocz161] [Medline: 31592533]
18. Moniz L, Buczak AL, Hung L, Babin S, Dorko M, Lombardo J. Construction and validation of synthetic electronic medical records. Online J Public Health Inform 2009;1(1) [FREE Full text] [doi: 10.5210/ojphi.v1i1.2720] [Medline: 23569572]
19. Loong B. DASH - Harvard University. 2012. Topics and Applications in Synthetic Data URL: https://dash.harvard.edu/bitstream/handle/1/9527319/Loong_gsas.harvard_0084L_10323.pdf?sequence=1 [accessed 2020-01-20]
20. Hochberg I. Insulin detemir use is associated with higher occurrence of hypoglycemia in hospitalized patients with hypoalbuminemia. Diabetes Care 2018 Apr;41(4):e44-e46. [doi: 10.2337/dc17-1957] [Medline: 29437697]
21. Gorelik Y, Yaseen H, Heyman SN, Khamaisi M. Negligible risk of acute renal failure among hospitalized patients after contrast-enhanced imaging with iodinated versus gadolinium-based agents. Invest Radiol 2019 May;54(5):312-318. [doi: 10.1097/RLI.0000000000000534] [Medline: 30480553]
22. Khoury J, Bahouth F, Stabholz Y, Elias A, Mashiach T, Aronson D, et al. Blood urea nitrogen variation upon admission and at discharge in patients with heart failure. ESC Heart Fail 2019 Aug;6(4):809-816 [FREE Full text] [doi: 10.1002/ehf2.12471] [Medline: 31199082]
23. Leendertse AJ, Egberts AC, Stoker LJ, van den Bemt PM, HARM Study Group. Frequency of and risk factors for preventable medication-related hospital admissions in the Netherlands. Arch Intern Med 2008 Sep 22;168(17):1890-1896. [doi: 10.1001/archinternmed.2008.3] [Medline: 18809816]
24. van der Hooft CS, Dieleman JP, Siemes C, Aarnoudse AL, Verhamme KM, Stricker BH, et al. Adverse drug reaction-related hospitalisations: a population-based cohort study. Pharmacoepidemiol Drug Saf 2008 Apr;17(4):365-371. [doi: 10.1002/pds.1565] [Medline: 18302300]
25. Kongkaew C, Hann M, Mandal J, Williams SD, Metcalfe D, Noyce PR, et al. Risk factors for hospital admissions associated with adverse drug events. Pharmacotherapy 2013 Aug;33(8):827-837. [doi: 10.1002/phar.1287] [Medline: 23686895]
26. Valgimigli M, Bueno H, Byrne RA, Collet JP, Costa F, Jeppsson A, ESC Scientific Document Group, ESC Committee for Practice Guidelines (CPG), ESC National Cardiac Societies. 2017 ESC focused update on dual antiplatelet therapy in coronary artery disease developed in collaboration with EACTS: The Task Force for dual antiplatelet therapy in coronary artery disease of the European Society of Cardiology (ESC) and of the European Association for Cardio-Thoracic Surgery (EACTS). Eur Heart J 2018 Jan 14;39(3):213-260. [doi: 10.1093/eurheartj/ehx419] [Medline: 28886622]
27. Bhatt DL, Scheiman J, Abraham NS, Antman EM, Chan FK, Furberg CD, American College of Cardiology Foundation Task Force on Clinical Expert Consensus Documents. ACCF/ACG/AHA 2008 expert consensus document on reducing the gastrointestinal risks of antiplatelet therapy and NSAID use: a report of the American College of Cardiology Foundation Task Force on Clinical Expert Consensus Documents. Circulation 2008 Oct 28;118(18):1894-1909. [doi: 10.1161/CIRCULATIONAHA.108.191087] [Medline: 18836135]
28. Anderson JL, Morrow DA. Acute myocardial infarction. N Engl J Med 2017;376:2053-2064. [doi: 10.1056/nejmra1606915]
29. McNamara RL, Wang Y, Herrin J, Curtis JP, Bradley EH, Magid DJ, NRMI Investigators. Effect of door-to-balloon time on mortality in patients with ST-segment elevation myocardial infarction. J Am Coll Cardiol 2006 Jun 6;47(11):2180-2186 [FREE Full text] [doi: 10.1016/j.jacc.2005.12.072] [Medline: 16750682]
30. Nallamothu BK, Normand ST, Wang Y, Hofer TP, Brush JE, Messenger JC, et al. Relation between door-to-balloon times and mortality after primary percutaneous coronary intervention over time: a retrospective study. Lancet 2015 Mar 21;385(9973):1114-1122 [FREE Full text] [doi: 10.1016/S0140-6736(14)61932-2] [Medline: 25467573]
31. Teerlink JR, Alburikan K, Metra M, Rodgers JE. Acute decompensated heart failure update. Curr Cardiol Rev 2015;11(1):53-62 [FREE Full text] [doi: 10.2174/1573403x09666131117174414] [Medline: 24251454]
32. Mehran R, Nikolsky E. Contrast-induced nephropathy: definition, epidemiology, and patients at risk. Kidney Int Suppl 2006 Apr(100):S11-S15 [FREE Full text] [doi: 10.1038/sj.ki.5000368] [Medline: 16612394]
33. Rao QA, Newhouse JH. Risk of nephropathy after intravenous administration of contrast material: a critical literature analysis. Radiology 2006 May;239(2):392-397. [doi: 10.1148/radiol.2392050413] [Medline: 16543592]
34. Hinson JS, Ehmann MR, Fine DM, Fishman EK, Toerper MF, Rothman RE, et al. Risk of acute kidney injury after intravenous contrast media administration. Ann Emerg Med 2017 May;69(5):577-86.e4. [doi: 10.1016/j.annemergmed.2016.11.021] [Medline: 28131489]

35. Umpierrez GE, Hellman R, Korytkowski MT, Kosiborod M, Maynard GA, Montori VM, Endocrine Society. Management of hyperglycemia in hospitalized patients in non-critical care setting: an endocrine society clinical practice guideline. J Clin Endocrinol Metab 2012 Jan;97(1):16-38. [doi: 10.1210/jc.2011-2098] [Medline: 22223765]
36. Goldman-Levine JD, Lee KW. Insulin detemir--a new basal insulin analog. Ann Pharmacother 2005 Mar;39(3):502-507. [doi: 10.1345/aph.1E334] [Medline: 15657117]
37. Reilly JB, Berns JS. Selection and dosing of medications for management of diabetes in patients with advanced kidney disease. Semin Dial 2010;23(2):163-168. [doi: 10.1111/j.1525-139X.2010.00703.x] [Medline: 20210915]

Abbreviations

ADHF: acute decompensated heart failure
AKI: acute kidney injury
BUN: blood urea nitrogen
CHF: congestive heart failure
CIN: contrast-induced nephropathy
CT: computed tomography
D2B: door-to-balloon time
DAT: double antiplatelet therapy
EMR: electronic medical record
IHD: ischemic heart disease
IRB: institutional review board
MRI: magnetic resonance imaging
OAC: oral anticoagulant
OSIM: Observational Medical Dataset Simulator
PCI: percutaneous coronary intervention
PPI: proton pump inhibitor
STEMI: ST-Elevation Myocardial Infarction

Edited by C Lovis; submitted 03.10.19; peer-reviewed by U Kartoun, M Westphal; comments to author 26.10.19; revised version received 01.12.19; accepted 27.12.19; published 20.02.20.

Please cite as:
Reiner Benaim A, Almog R, Gorelik Y, Hochberg I, Nassar L, Mashiach T, Khamaisi M, Lurie Y, Azzam ZS, Khoury J, Kurnik D, Beyar R
Analyzing Medical Research Results Based on Synthetic Data and Their Relation to Real Data Results: Systematic Comparison From Five Observational Studies
JMIR Med Inform 2020;8(2):e16492
URL: http://medinform.jmir.org/2020/2/e16492/
doi:10.2196/16492
PMID:32130148

Original Paper

Detection of Postictal Generalized Electroencephalogram Suppression: Random Forest Approach

Xiaojin Li[1], PhD; Shiqiang Tao[1], PhD; Shirin Jamal-Omidi[1], MD; Yan Huang[2], MSc; Samden D Lhatoo[1], MD; Guo-Qiang Zhang[1,3], PhD; Licong Cui[3], PhD

[1]Department of Neurology, University of Texas Health Science Center, Houston, TX, United States

[2]Department of Computer Science, University of Kentucky, Lexington, KY, United States

[3]School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX, United States

Corresponding Author:
Licong Cui, PhD
School of Biomedical Informatics
University of Texas Health Science Center
7000 Fannin St
Houston, TX, 77030
United States
Phone: 1 7135003791
Email: licong.cui@uth.tmc.edu

Abstract

Background: Sudden unexpected death in epilepsy (SUDEP) is second only to stroke in neurological events resulting in years of potential life lost. Postictal generalized electroencephalogram (EEG) suppression (PGES) is a period of suppressed brain activity often occurring after generalized tonic-clonic seizure, a most significant risk factor for SUDEP. Therefore, PGES has been considered as a potential biomarker for SUDEP risk. Automatic PGES detection tools can address the limitations of labor-intensive, and sometimes inconsistent, visual analysis. A successful approach to automatic PGES detection must overcome computational challenges involved in the detection of subtle amplitude changes in EEG recordings, which may contain physiological and acquisition artifacts.

Objective: This study aimed to present a random forest approach for automatic PGES detection using multichannel human EEG recordings acquired in epilepsy monitoring units.

Methods: We used a combination of temporal, frequency, wavelet, and interchannel correlation features derived from EEG signals to train a random forest classifier. We also constructed and applied confidence-based correction rules based on PGES state changes. Motivated by practical utility, we introduced a new, time distance–based evaluation method for assessing the performance of PGES detection algorithms.

Results: The time distance–based evaluation showed that our approach achieved a 5-second tolerance-based positive prediction rate of 0.95 for artifact-free signals. For signals with different artifact levels, our prediction rates varied from 0.68 to 0.81.

Conclusions: We introduced a feature-based, random forest approach for automatic PGES detection using multichannel EEG recordings. Our approach achieved increasingly better time distance–based performance with reduced signal artifact levels. Further study is needed for PGES detection algorithms to perform well irrespective of the levels of signal artifacts.

(JMIR Med Inform 2020;8(2):e17061) doi:10.2196/17061

KEYWORDS

epilepsy; generalized tonic-clonic seizure; postictal generalized EEG suppression; EEG; random forest

Introduction

Background

Epilepsy is one of the most common neurological disorders, and it affects an estimated 65 million people worldwide [1]. An epileptic seizure (hereafter referred to as seizure) is a brief episode, usually with signs or symptoms because of transient, undesired, excessive, and synchronous electrical discharge, involving large numbers of neurons in the brain [2]. When seizure occurs, altered movement, expression, and levels of consciousness are often observed in the affected person. Seizure

may produce temporary confusion, uncontrollable jerking movements of the arms and legs, inability to speak, or loss of consciousness or awareness [3].

In a worst-case scenario, frequent seizures may predispose a person to sudden unexpected death in epilepsy (SUDEP) [4]. Among neurological events and conditions, SUDEP is second only to stroke in years of potential life lost, highlighting the importance and significance of this condition for public health [5]. SUDEP is a catastrophic and fatal complication of epilepsy. The definition of SUDEP is "sudden, unexpected, witnessed or unwitnessed, non-traumatic and non-drowning death, occurring in benign circumstances, in an individual with epilepsy, with or without evidence for a seizure and excluding documented status epilepticus, in which postmortem examination does not reveal a cause of death" [6], that is, no other cause of death can be found [7]. However, the mechanisms underlying SUDEP are not completely understood.

Electrophysiological signals such as electroencephalogram (EEG), electrocardiogram, and electromyography, collected together in the epilepsy monitoring unit (EMU), are traditionally used for understanding epileptic seizures [8]. Noninvasive scalp EEG and invasive intracranial EEG are the most commonly used methods for locating seizures and monitoring the interphase activity between seizures [9]. Invasive intracranial EEG is one of the techniques used in localizing the seizure onset zone in preparation for surgery [8]. EEG is a key source of information for the diagnosis of epilepsy, including whether epilepsy is focal or generalized, idiopathic or symptomatic, or part of a specific epilepsy syndrome [10]. Therefore, EEG has also been widely used to identify biomarkers that can help prevent the development of epilepsy, identify focal brain regions that produce epilepsy, and ultimately cure epilepsy through surgical means [11].

Postictal generalized EEG suppression (PGES) is a potential EEG biomarker of SUDEP risk [12-14]. PGES is a period of brain inactivity after seizure. It most often occurs after generalized tonic-clonic seizures (GTCS), particularly in those arising from sleep, and is related to the symmetric tonic phase, postictal immobility, lack of early oxygen administration, duration of oxygen desaturation, and lower peripheral capillary oxygen saturation nadir values [15-17]. GTCS are the most significant risk factor for SUDEP [13]. PGES is defined as a diffused EEG background attenuation (<10 μV) in the postictal period [18]. Prolonged PGES (>50 seconds) has been reported in refractory epilepsy patients who are at risk of SUDEP [14]. For every prolonged second in the duration of PGES, the odds of SUDEP is increased by a factor of 1.7% (P<.005) [14].

Clinically, the determination of the duration of PGES is manually performed by human experts through visual inspection of EEG signals. According to definition, the identification of PGES appears to be straightforward by identifying a period of low-amplitude EEG signals after the seizure, as shown in Figure 1. However, real-world data recorded in EMUs may contain high-amplitude signals caused by physiological artifacts (eg, breathing, muscle, and movement artifacts), as shown in Figure 2. Therefore, clinical experts usually leverage additional video recordings along with signals to identify high-amplitude artifacts that are not real EEG activities [19]. Automated PGES detection tools are highly desirable to assist clinical personnel in reviewing and annotating PGES in EEG recordings. Automated techniques have been extensively studied for epilepsy-related EEG signal analysis [20], including a random forest classifier with empirical wavelet transform for seizure identification [21], a data-driven approach for classifying seizure and nonseizure EEG signals using the multivariate empirical mode decomposition algorithm [22], a whole-brain seizure detection approach using the K-nearest neighbors classifier [23], and extreme epileptic events detection and prediction using neural networks with time-frequency features [24,25]. However, there has been only one study [19] using logistic regression to perform automated PGES detection based on frequency-domain features of EEG signals. The following challenges remain in developing a fully automatic PGES detection tool:

- The presence of artifacts remains the main challenge that makes PGES detection more complex than applying a fixed amplitude threshold.
- There is no sensitive and standardized criterion dedicated to measuring and evaluating the performance of PGES detection algorithms.

Figure 1. An example of postictal generalized electroencephalogram suppression and intermittent slow wave activity signals after generalized tonic-clonic seizures. GTCS: generalized tonic-clonic seizures; ISW: intermittent slow wave; PGES: postictal generalized electroencephalogram suppression.

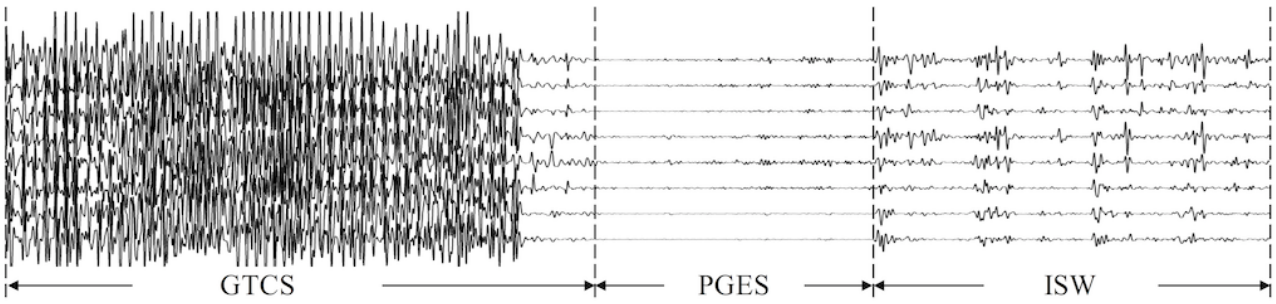
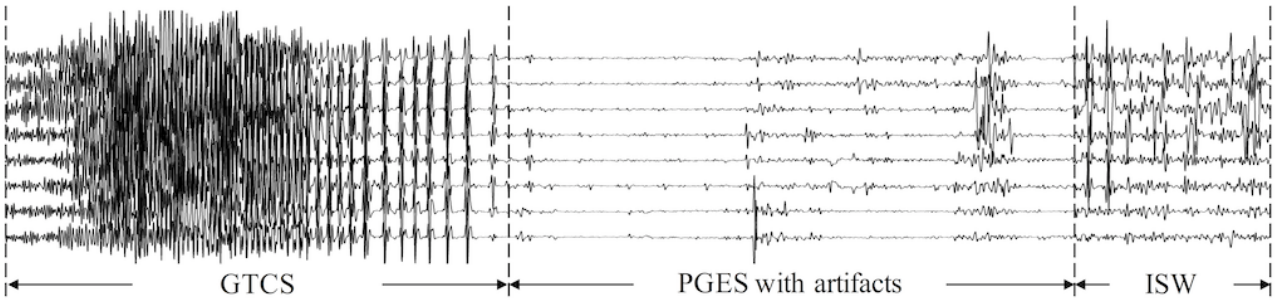


Figure 2. An example of postictal generalized electroencephalogram suppression and intermittent slow wave activity signals (with artifacts) after generalized tonic-clonic seizures. GTCS: generalized tonic-clonic seizures; ISW: intermittent slow wave; PGES: postictal generalized electroencephalogram suppression.



In this paper, we introduce a random forest–based classifier for PGES detection by leveraging a variety of EEG signal features such as time-domain features, frequency-domain features, wavelet-based features, and interchannel correlations. We incorporate confidence-based correction rules to remove suspicious sudden changes of EEG activities. This study focused on identifying the first slow wave brain activity, that is, the onset of the first intermittent slow wave (ISW) activity (see Figure 1 and Figure 2), which indicates that the brain activity will gradually recover [12,14]. Therefore, the output of our PGES detection method for each signal recording is the onset of the first ISW. Accordingly, traditional segment-based performance evaluation methods are not well suited for PGES detection. Instead, we introduced a new, recording-by-recoding evaluation method dedicated to PGES detection with direct practical relevance.

The Center for Sudden Unexpected Death in Epilepsy Research

The Center for SUDEP Research (CSR) is a National Institute of Neurological Disorders and Stroke–funded Center Without Walls initiative for collaborative research on epilepsy. It comprises researchers from 14 institutions across the United States and Europe, bringing together extensive and diverse expertise to understand SUDEP [4,26]. The goal of CSR is to better understand cortical, subcortical, and brainstem mechanisms responsible for SUDEP and to use a data-driven, systems biology approach to elucidate the role of cortical influences in SUDEP. To advance SUDEP research, CSR created an infrastructure to fully, effectively, and efficiently utilize a range of prospectively collected data from different domains, including clinical, electrophysiological, biochemical, genetic, and neuropathological fields. CSR provides a comprehensive, curated repository of prospectively collected multimodal data, including electrophysiological signals in European data format. These data are linked to risk factor and outcomes data of over 2500 epilepsy patients (a broad spectrum of ages as well as social, racial, and ethnic groups) with thousands of 24-hour recordings.

Feature Extraction From Electroencephalogram Signals

For EEG signal feature extraction, the following 4 categories of features are considered in this work: (1) time-domain features, (2) frequency-domain features, (3) wavelet-based features, and (4) interchannel correlations.

1. Time-domain features: Time-domain features include statistical measures and Hjorth parameters. Statistical measures include nth percentile of the signal, average, range, standard deviation, skewness, and kurtosis [27]. Here, the mean measures the central tendency, skewness measures the asymmetry, and kurtosis measures the tailedness of a probability distribution. Hjorth parameters are commonly used for feature extraction to perform EEG signal analysis [28], including mobility and complexity [29-31]. The mobility represents the mean frequency or the proportion of the standard deviation of the power spectrum. The complexity indicates the signal's similarity to a pure sine wave [31].
2. Frequency-domain features: An EEG wave (captured by an electrode) comprises many other waves with different amplitudes and frequencies. Therefore, an EEG signal has different bands, defined by the frequency of the waves, such as slow oscillations (0.5 Hz-1 Hz), delta bands (1 Hz-4 Hz), theta bands (4 Hz-8 Hz), alpha bands (8 Hz-12 Hz), beta bands (14 Hz-30 Hz), and gamma bands (30 Hz-80 Hz). The spectral power in a specific frequency band, for instance, the 0.5 Hz to 1 Hz band, can be regarded as a feature.
3. Wavelet-based features: Wavelets are a relatively recent approach for signal processing, and the main advantage is that wavelets allow multiresolution analysis in time and frequency simultaneously [32,33].
4. Interchannel correlations: Many studies have attempted to find movement-related information of connectivity between different brain regions [34-38]. Correlation analysis represents the degree of relatedness and synchrony between 2 time series. It indicates similar information as a cross-coherence analysis of different EEG channels [39].

Random Forest

Random forest is an ensemble learning method used for classification and regression problems. This method has been used for automated sleep stage classification based on EEG signals [40,41]. Random forest involves a group of decision trees during training and outputs the mode of the classes predicted by individual trees. The overall output is determined by applying the input to each tree and choosing the class that gets the most weighted vote. The weight of each tree is adjusted using misclassification and out-of-bag measures.

As it is different from other traditional classifiers (eg, K-nearest neighbor, support vector machine, and artificial neural network),

we select random forest as the approach for our study because of the following advantages: (1) adaptable, as it estimates the importance of variables and provides a way for tuning with additional training data by assigning different weights for each decision tree; (2) scalable, as it can handle thousands of input variables and work efficiently on large datasets; and (3) robust, as it can balance errors in datasets with unbalanced class population [41].

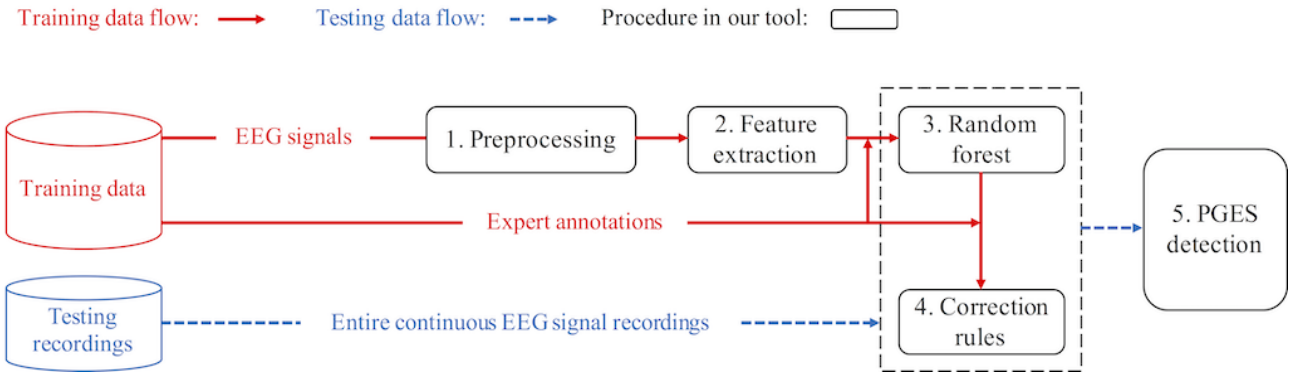
Methods

Overview

The dataset used for this study comprises 116 EEG signal recordings from 84 patients with GTCS in the CSR data repository, with PGES annotated by domain experts. We extracted the 5-min postictal EEG signals for PGES detection. A total of 8 EEG channels are utilized: Fp1-F7, F7-T7, T7-P7, Fp2-F8, F8-T8, T8-P8, Fz-Cz, and Cz-Pz.

The overall workflow of our PGES detection method is shown in Figure 3. The process started with the preprocessing of the EEG signals (step 1), followed by feature extraction (step 2). Then a random forest classifier was trained and tested based on the extracted features (step 3). We applied correction rules, which are constructed based on the continuity of brain activities, to the prediction of the random forest (step 4) and provided the final detected label for each signal segment (step 5).

Figure 3. The overall workflow of our automated postictal generalized electroencephalogram suppression detection approach. EEG: electroencephalogram; PGES: postictal generalized electroencephalogram suppression.



Preprocessing

Each postictal EEG signal record is split into signal segments with a length of 1 second (ie, 1-second epoch) from the beginning to the end without overlapping. The common electrophysiological artifacts present in the EEG signal recordings include muscle artifacts, breathing, and body and bed movements [42]. The main frequency of ISW is less than 5 Hz. To minimize the presence of residual artifacts, the signals are filtered with a band-pass filter with cutoff frequencies at 0.5 Hz and 5 Hz.

Feature Extraction

For each signal segment of the 8 EEG channels, we extracted 76 features including time-domain features, frequency-domain features, and wavelet-based features as follows:

- The following 16 time-domain features were extracted: (1) 11 statistic features, including mean, median, maximum, minimum, range, standard deviation, *n*th percentile of the signal (n=5, 25, 75, and 95), and the root mean square, and (2) 5 time-domain properties of kurtosis, skewness, mobility, complexity, and amplitude energy (AE) of the signal. The time-domain properties for a time series $X=\{x_1, x_2, ..., x_n\}$ are defined in Figure 4 [27-30], where N is the number of data points, $\bar{x}$ is the mean of X, and $d_i=x_i-x_{i-1}$ *and* $i=1, ..., n$.
- A total of 4 frequency-domain features were extracted. As the PGES and ISW are typically low-frequency brain activities (0.5 Hz-5 Hz), we extracted the spectral power of 4 low-frequency bands consisting of 0.5 Hz to 1 Hz, 1 Hz to 2 Hz, 2 Hz to 4 Hz, and 4 Hz to 5 Hz.
- A total of 56 wavelet-based features were extracted. EEG signals were subjected to three-level decomposition using the Daubechies 4 wavelet. From the decomposition process, a total of 4 coefficient sets were generated, and we calculated 14 measurements (except range and AE) used in time-domain features for each coefficient set as wavelet-based features [32,33].

Figure 4. The definitions of the time-domain properties. AE: amplitude energy.

$$Kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^4}{[\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2]^2} - 3 \text{ (a)}$$

$$Skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^3}{[\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2]^{3/2}} \text{ (b)}$$

$$Mobility = \sqrt{\frac{\sum_{i=1}^{N} d_i}{\sum_{i=1}^{N} x_i}} \text{ (c)}$$

$$Complexity = \sqrt{\frac{[\sum_{i=1}^{N}(d_i-d_{i-1})^2]\times\sum_{i=1}^{N} x_i}{(\sum_{i=1}^{N} d_i)^2}} \text{ (d)}$$

$$AE = \sum_{i=1}^{N}|x_i| \text{ (e)}$$

To capture cross-coherence of EEG channels, we further investigated the interchannel correlations using the linear correlation coefficient between selected channels. The correlation coefficient for 2 time series, $X=\{x_1, x_2, ..., x_n\}$ and $Y=\{y_1, y_2, ..., y_n\}$ is defined in Figure 5 [39], where N is the number of data points and $\bar{x}$ is the mean. For each signal segment, we calculated 4 interchannel correlations: *corr(Fp1−F7, Fp2−F8)*, *corr(Fp2−F8, Fz−Cz)*, *corr(Fp1−F7, Fz−Cz)*, and *corr(Fz−Cz, Cz−Pz)*.

Figure 5. The definitions of the linear correlation coefficient between two time series (X and Y).

$$corr(X,Y) = \frac{\sum_{i=1}^{N}(x_i-\bar{x})\times(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i-\bar{y})^2}}$$

Random Forest Classifier

There are 5 steps to build the random forest classifier with the bootstrap aggregating (bagging) technique [43], which is an ensemble method to reduce the variance without increasing the bias for decision tree algorithms. Given a training set $X=\{x_1, x_2, ..., x_n\}$ and $Y=\{y_1, y_2, ..., y_n\}$, the 5 steps are as follows:

1. Generate a subtraining set $S=\{X_s, Y_s\}$ by selecting a random number of observations and features from the whole input training dataset.
2. Build and train each decision tree T_i (eg, regression tree, random tree, and C4.5) with the generated subtraining set S. In the construction of each decision tree, nodes and leaves are built by selecting a random number of features. This process will minimize the correlation among the features and decrease the sensitivity to noise [40].
3. Estimate out-of-bag errors. In the training process of each tree, about two-thirds of S are used for tree construction, and the remaining one-third is used to test the classification performance of the tree. Therefore, it gets an unbiased estimate of the test set error internally in a random forest, and there is no need to use further cross-validation [43].
4. Repeat the above steps (step 1, step 2, and step 3) N times to build N decision trees $T=\{T_1, T_2, ..., T_N\}$.
5. Compute the classifier output. After training, the output y' for an unknown sample x' can be made by averaging the output from all the individual decision trees on x':

This classifier can obtain strongly correlated trees by training all trees with the same training set, and bagging is a way to decorrelate the trees. The prediction results of a single decision tree may be highly sensitive to noises in its particular training set, especially with overfitting. However, in a random forest, the average of all trees is less sensitive to noises as the trees are more decorrelated. In this work, we used a random forest with $N=1000$ trees.

Correction Rules for Continuous Detection

According to the knowledge of the domain experts, longtime EEG suppression does not often occur after the first ISW happens in practical scenarios. This indicates that sudden changes of PGES/ISW states are unlikely to happen. For example, for a sequence of predicted labels with 10 consecutive 1-second segments (PGES, PGES, PGES, ISW, PGES, PGES, PGES, PGES, PGES, and PGES), the sudden changes from PGES to ISW (from the third segment to the fourth segment) and from ISW back to PGES (from the fourth segment to the fifth segment) are unlikely, that is, the predicted label for the fourth segment is most likely a misclassification and should be corrected and replaced with PGES. Therefore, we considered

the temporal contextual information of segments to perform correction.

We constructed confidence-based correction rules based on the probability output of the random forest classifier. For each segment Seg_i, we built a confidence index, $conf(Seg_i)$ based on the average probability from the current segment to the next M segments; the definition is as follows:

where $prob(Seg_j)$ is the random forest probability output of the segment Seg_j.

In this work, we chose $M=5$. On the basis of the probability and confidence index of each segment, we applied the following 3 rules to correct suspicious sudden changes of PGES/ISW states in the detected label sequence:

1. If $prob(Seg_i)$ had a high value but the probabilities of Seg_i's surrounding segments (eg, $prob(Seg_{i-1})$ and $prob(Seg_{i+1})$ had low values, then we corrected the detected label of *Segi* as PGES.
2. If $conf(Seg_i)$ had a high value but the confidence indexes of Seg_i's surrounding segments (eg, $conf(Seg_{i-1})$ and $conf(Seg_{i+1})$) had low values, then we corrected the detected label of Seg_i as PGES.
3. If $prob(Seg_i)$ had a high value but $conf(Seg_i)$ had a low value, then we corrected the detected label of Seg_i as PGES.

Performance Evaluation

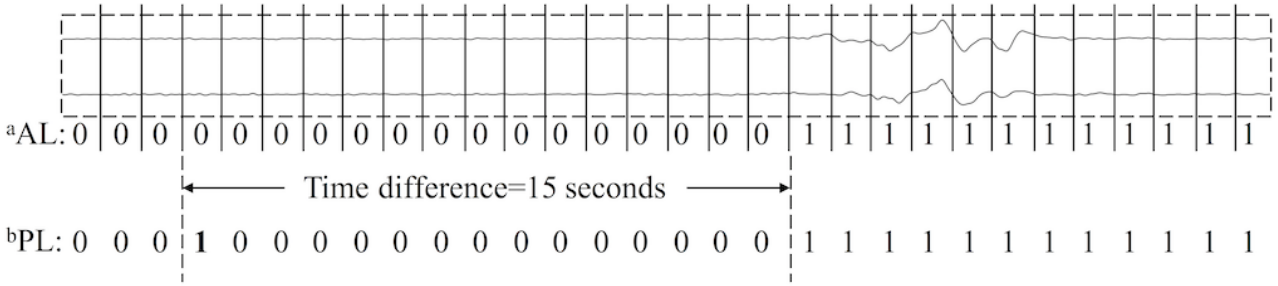
For traditional EEG signal classification tasks such as sleep stage classification [40,44,45], the performance evaluations are segment-based (ie, the predictions for each segment determine the performance metrics such as accuracy, precision, and recall). However, for the PGES detection setting, the prediction result of the onset of the first ISW in a given signal recording (ie, recording-based) is more important as it indicates the end of PGES. In other words, the prediction results of segments after the first ISW become less important for the evaluation. Figure 6 shows an example of the signal that is split into 30 segments of 1-second each, and each segment is annotated with a label *0* for PGES or *1* for ISW (see the annotated labels). The predicted labels are generated by the automatic detection method; the predicted label highlighted in bold means that the label is wrongly predicted, whereas other labels are correctly predicted. In this example, only one label is wrongly predicted, indicating that the PGES detection method achieves a high accuracy of 97% (29/30) for the segment-based evaluation. However, the actual first ISW of the signal is 15 seconds away from the predicted first ISW, a 15-second time difference that may not be acceptable in clinical scenarios.

Therefore, for the first time, we proposed time distance–based metrics to evaluate an automated PGES detection method. For a given signal recording r, we defined the *predicted time distance* TD_r as the time difference between the predicted end time of PGES (or the predicted time of the first ISW) by the detection method and the actual end time of PGES (or the actual time of the first ISW) according to the expert annotations. A lower value of the time distance indicates a better performance of the PGES detection method.

On the basis of the predicted time distance, we further introduced the 5-second tolerance-based positive prediction rate (PPR_{5s}) as another evaluation metric, as a 5-second time distance is acceptable according to clinical experts. Given a collection R of signal recordings for evaluation, we define PPR_{5s} as follows:

that is, the number of signal recordings whose predicted time distances are within 5 seconds divided by the total number of the signal recordings.

Figure 6. An example of automatic postictal generalized electroencephalogram suppression detection evaluation. Annotated labels are the expert-annotated labels, and predicted labels are generated using the automatic detection algorithm. [a]AL: annotated label; [b]PL: predicted label.



Results

Artifact Level

To evaluate the performance of our PGES detection method on EEG signals with different levels of artifacts, we categorized EEG signals into 4 levels: artifact-free, mild artifact, moderate artifact, and severe artifact. The domain expert manually reviewed the EEG signals and classified them into 4 levels according to the following criteria:

1. *Artifact-free*: No waveforms in any channels or abrupt waveforms of less than a second duration.
2. *Mild*: Abrupt waveforms in channels other than midline channels (Fz-Cz and Cz-Pz) that do not affect the midline channels or midline channels involved with abrupt waveforms of less than 1-second duration.

3. *Moderate*: One of the midline channels is involved, or both midline channels are involved with waveforms of less than 1-second duration.
4. *Severe*: Both midline channels have abrupt waveforms of more than 1-second duration. An expert may need to analyze all EEG chains, which include 19 EEG channels, or use video recordings to differentiate the artifacts from the brain-generated waveforms.

Among the 116 signal recordings in our dataset, 27 are artifact-free, 31 are with mild artifacts, 25 are with moderate artifacts, and 33 are with severe artifacts. We applied our PGES detection method to 4 groups of signal recordings with different levels of artifacts: only artifact-free (group A); artifact-free and mild artifact (group B); artifact-free, mild artifact, and moderate artifact (group C); and all signal recordings (group D).

Cross-Validation

Cross-validation has been generally used for evaluating a model's performance with low bias and variance. We applied 10-fold cross-validation 10 times to the 4 groups with varying artifact levels to evaluate our PGES detection method. For each group, we randomly separated the signal recordings into two parts each time, training set and testing set, and then calculated the evaluation metrics. Note that there was no overlap between the training set and testing set. For instance, in group D, which included all 116 signal recordings, 11 recordings were used as the testing data and the remaining 105 recordings were used as the training data in each fold. We repeated this procedure 10 times and used the average as the final evaluation result.

For the training set, we selected balanced numbers of PGES and ISW segments for each signal recording. Although every EEG recording was annotated by domain experts with the start of the first ISW, there existed EEG recordings missing the annotations of the end of the first ISW (as the onset of the first ISW is the most important). Therefore, for ISW signal segments, we selected up to 30-second signal (ie, 30 segments) after the onset of the first ISW as follows: if the duration of PGES (say t seconds) is less than 30 seconds, then we used t segments after the first onset of ISW; otherwise, we used 30 segments after the first onset of ISW. For the testing set, we used the entire 5 min of each signal recording to detect the first onset of ISW.

The average predicted time distance is 2.4 seconds for artifact-free signal recordings (group A) and 4.34 seconds for the group containing both artifact-free and mild artifact signal recordings (group B). As the artifact level increases, the average predicted time distance increases as well. The average predicted time distance is 7.54 seconds for the group containing artifact-free, mild artifact, and moderate artifact signal recordings (group C) and 7.84 seconds for all signal recordings (group D).

The PPR_{5s} of our PGES detection method for each artifact group of signal recordings is as follows: PPR_{5s} is 0.95 for artifact-free signal recordings (group A); 0.81 for the group containing both artifact-free and mild artifact signal recordings (group B); 0.73 for the group containing artifact-free, mild artifact, and moderate artifact signal recordings (group C); and 0.68 for all signal recordings (group D). It can be seen that PPR_{5s} decreases as the level of artifacts increases.

As a comparison, we also calculated the segment-based evaluation metrics to evaluate the performance of our PGES detection method for classifying individual signal segments, including accuracy, recall (R_{PGES}), precision (P_{PGES}), and F1-score ($F1_{PGES}$), as defined in Figure 7, where TP_{PGES} is the number of segments detected as PGES and labeled as PGES by experts, TP_{PGES} is the number of segments detected as ISW and labeled as ISW by experts, Num_{PGES_expert} is the total number of segments labeled as PGES by experts, Num_{PGES_detect} is the total number of segments labeled as PGES by our detection method, and Num_{total} is the total number of segments.

Table 1 shows the segment-based evaluation results of our method. The accuracy of each artifact group is over 0.92. On the other hand, the recall, precision, and F1-score for each group are over 0.95, 0.96, and 0.95, respectively.

Figure 7. The definitions of segment-based evaluation metrics.

$$Accuracy = \frac{TP_{PGES}+TN_{PGES}}{Num_{total}} \text{ (a)}$$

$$Recall = \frac{TP_{PGES}}{Num_{PGES\_expert}} \text{ (b)}$$

$$Precision = \frac{TP_{PGES}}{Num_{PGES\_detect}} \text{ (c)}$$

$$F1\text{-}score = 2\times\frac{Recall\times Precision}{Recall+Precision} \text{ (d)}$$

Table 1. The traditional evaluation of postictal generalized electroencephalogram suppression (PGES) detection on each testing group.

Evaluation metric	Group A[a]	Group B[b]	Group C[c]	Group D[d]
Accuracy	0.94	0.94	0.94	0.92
Recall (R_{PGES})	0.95	0.96	0.97	0.95
Precision (P_{PGES})	0.98	0.98	0.97	0.96
F1-score ($F1_{PGES}$)	0.97	0.96	0.97	0.95

[a]Only artifact-free signal recordings.

[b]Artifact-free and mild artifact signal recordings.

[c]Artifact-free, mild artifact, and moderate artifact signal recordings.

[d]All signal recordings.

Discussion

Principal Findings

We developed an automated PGES detection method based on EEG signals, which combined a random forest classifier and 3 correction rules. The main idea of our method was to leverage both signal features and the state transitions of brain activities. We evaluated the performance of our method using different artifact groups of signal recordings.

We reported both the segment-based evaluation results and the recording-based evaluation results. According to the segment-based evaluation results, our method achieved over 0.92 accuracy, 0.95 recall, 0.96 precision, and 0.95 F1-score for each artifact group. The results were consistent for each group, which indicates that our method performed well for classifying individual PGES signal segments (even for the group containing signal recording with severe artifacts).

However, as illustrated in Figure 6, the segment-based evaluation may not be able to demonstrate the actual PGES detection performance. In practical settings, the segment that was incorrectly detected as ISW would cause the wrong annotation of the PGES end time and result in an incorrect, significantly different PGES duration, which may mislead the risk assessment of SUDEP. Therefore, we introduced a way with direct practical relevance to evaluate automated PGES methods based on time distance, which is the time difference between the detected PGES period and the expert-annotated one.

On the basis of our recording and time distance–based evaluation, our PGES detection method achieved an average predicted time distance of 2.4 seconds and a PPR_{5s} of 0.95 for artifact-free EEG signals. For signals with artifacts, the performance of this method varies according to the level of artifacts. For signals with mild artifacts, our method achieved a PPR_{5s} of 0.81. However, as the number of signals with higher artifact levels (moderate) increased, the PPR_{5s} dropped to 0.73; for signals with all artifact levels (artifact-free to severe), it dropped to 0.68, and the average predicted time distance was 7.84 seconds. The artifact is the main challenge for PGES detection. To identify high-amplitude artifacts (severe level) that are not real brain activities, clinicians usually have to use different EEG patterns or even video recordings. In future work, we will focus on developing an approach for handling signals with high artifact levels (moderate and above). In particular, we plan to try dedicated artifact removal methods such as independent component analysis [46,47], regression analysis [48], and empirical method [49,50] to study whether such methods would help improving the performance of PGES detection.

Compared with the previous work [19], we used three additional types of features (time-domain features, wavelet-based features, and interchannel correlations) in the feature extraction step. For the classifier, we used random forest instead of boosting algorithms with logistic regression. We also introduced a new metric (*predicted time distance*) to evaluate an automated PGES detection method, and we reported evaluation results for both the segment-based method and our new metrics (no such evaluations were performed in the study by Theeranaew et al [19]).

Conclusions

We presented an automated method that combines the benefits of random forest classifier and correction rules for PGES detection using multichannel EEG recordings. Features from temporal, frequency, wavelet, and cross-coherence analyses provided valuable information to characterize PGES and ISW. Confidence-based rules were leveraged to correct sudden changes of PGES states. We introduced a new evaluation method for assessing PGES detection performance with more practical relevance. The evaluation results indicated that our method achieved a PPR_{5s} of 0.95 for artifact-free EEG recordings. For EEG recordings with different artifact levels, the PPR_{5s} varied from 0.68 to 0.81. This study demonstrates that our combined random forest and rule-based approach can perform well in realistic settings for good quality EEG recordings.

Acknowledgments

This work was supported by the US National Institutes of Health (NIH) under grant U01NS090408. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Conflicts of Interest

None declared.

References

1. Bentivoglio M, Cavalheiro EA, Kristensson K, Patel NB, editors. Neglected Tropical Diseases and Conditions of the Nervous System. Heidelberg: Springer; 2014.
2. Goldenberg MM. Overview of drugs used for epilepsy and seizures: etiology, diagnosis, and treatment. Pharm Ther 2010 Jul;35(7):392-415 [FREE Full text] [Medline: 20689626]
3. Good DC. Chapter 51. Episodic neurologic symptoms. In: Walker HK, Hall WD, Hurst JW, editors. Clinical Methods: The History, Physical and Laboratory Examinations. Third Edition. Boston: Butterworths; 1990.
4. Lhatoo SD, Nei M, Raghavan M, Sperling M, Zonjy B, Lacuey N, et al. Nonseizure SUDEP: Sudden unexpected death in epilepsy without preceding epileptic seizures. Epilepsia 2016 Jul;57(7):1161-1168 [FREE Full text] [doi: 10.1111/epi.13419] [Medline: 27221596]
5. Thurman DJ, Hesdorffer DC, French JA. Sudden unexpected death in epilepsy: assessing the public health burden. Epilepsia 2014 Oct;55(10):1479-1485 [FREE Full text] [doi: 10.1111/epi.12666] [Medline: 24903551]
6. Maguire M, Jackson C, Marson A, Nolan SJ. Treatments for the prevention of sudden unexpected death in epilepsy (SUDEP). Cochrane Database Syst Rev 2016 Jul 19;7:CD011792 [FREE Full text] [doi: 10.1002/14651858.CD011792.pub2] [Medline: 27434597]
7. Sperling MR. Sudden unexplained death in epilepsy. Epilepsy Curr 2001 Sep;1(1):21-23 [FREE Full text] [doi: 10.1046/j.1535-7597.2001.00012.x] [Medline: 15309034]
8. Bertram EH. Electrophysiology in epilepsy surgery: roles and limitations. Ann Indian Acad Neurol 2014 Mar;17(Suppl 1):S40-S44 [FREE Full text] [doi: 10.4103/0972-2327.128649] [Medline: 24791088]
9. Worrell G, Gotman J. High-frequency oscillations and other electrophysiological biomarkers of epilepsy: clinical studies. Biomark Med 2011 Oct;5(5):557-566 [FREE Full text] [doi: 10.2217/bmm.11.74] [Medline: 22003904]
10. Smith SJ. EEG in the diagnosis, classification, and management of patients with epilepsy. J Neurol Neurosurg Psychiatry 2005 Jun;76(Suppl 2):ii2-ii7 [FREE Full text] [doi: 10.1136/jnnp.2005.069245] [Medline: 15961864]
11. Staba RJ, Stead M, Worrell GA. Electrophysiological biomarkers of epilepsy. Neurotherapeutics 2014 Apr;11(2):334-346 [FREE Full text] [doi: 10.1007/s13311-014-0259-0] [Medline: 24519238]
12. Vilella L, Lacuey N, Hampson JP, Rani MR, Loparo K, Sainju RK, et al. Incidence, recurrence, and risk factors for peri-ictal central apnea and sudden unexpected death in epilepsy. Front Neurol 2019;10:166 [FREE Full text] [doi: 10.3389/fneur.2019.00166] [Medline: 30890997]
13. Wu S, Issa NP, Rose SL, Ali A, Tao JX. Impact of periictal nurse interventions on postictal generalized EEG suppression in generalized convulsive seizures. Epilepsy Behav 2016 May;58:22-25. [doi: 10.1016/j.yebeh.2016.02.025] [Medline: 26994879]
14. Lhatoo SD, Faulkner HJ, Dembny K, Trippick K, Johnson C, Bird JM. An electroclinical case-control study of sudden unexpected death in epilepsy. Ann Neurol 2010 Dec;68(6):787-796. [doi: 10.1002/ana.22101] [Medline: 20882604]
15. Esmaeili B, Kaffashi F, Theeranaew W, Dabir A, Lhatoo SD, Loparo KA. Post-ictal modulation of baroreflex sensitivity in patients with intractable epilepsy. Front Neurol 2018;9:793 [FREE Full text] [doi: 10.3389/fneur.2018.00793] [Medline: 30319527]
16. Alexandre V, Mercedes B, Valton L, Maillard L, Bartolomei F, Szurhaj W, REPO2MSE study group. Risk factors of postictal generalized EEG suppression in generalized convulsive seizures. Neurology 2015 Nov 3;85(18):1598-1603. [doi: 10.1212/WNL.0000000000001949] [Medline: 26333799]
17. Kuo J, Zhao W, Li C, Kennedy JD, Seyal M. Postictal immobility and generalized EEG suppression are associated with the severity of respiratory dysfunction. Epilepsia 2016 Mar;57(3):412-417 [FREE Full text] [doi: 10.1111/epi.13312] [Medline: 26763069]
18. Asadollahi M, Noorbakhsh M, Simani L, Ramezani M, Gharagozli K. Two predictors of postictal generalized EEG suppression: tonic phase duration and postictal immobility period. Seizure 2018 Oct;61:135-138. [doi: 10.1016/j.seizure.2018.08.009] [Medline: 30142618]
19. Theeranaew W, McDonald J, Zonjy B, Kaffashi F, Moseley BD, Friedman D, et al. Automated detection of postictal generalized EEG suppression. IEEE Trans Biomed Eng 2018 Feb;65(2):371-377 [FREE Full text] [doi: 10.1109/TBME.2017.2771468] [Medline: 29346105]
20. Baumgartner C, Koren JP, Rothmayer M. Automatic computer-based detection of epileptic seizures. Front Neurol 2018;9:639 [FREE Full text] [doi: 10.3389/fneur.2018.00639] [Medline: 30140254]
21. Bhattacharyya A, Pachori RB. A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform. IEEE Trans Biomed Eng 2017;64(9):2003-2015. [doi: 10.1109/tbme.2017.2650259]
22. Zahra A, Kanwal N, Rehman NU, Ehsan S, McDonald-Maier KD. Seizure detection from EEG signals using Multivariate Empirical Mode Decomposition. Comput Biol Med 2017 Sep 1;88:132-141. [doi: 10.1016/j.compbiomed.2017.07.010] [Medline: 28719805]

23. Fergus P, Hussain A, Hignett D, Al-Jumeily D, Abdel-Aziz K, Hamdan H. A machine learning system for automated whole-brain seizure detection. Appl Comput Inform 2016;12(1):70-89. [doi: 10.1016/j.aci.2015.01.001]
24. Pisarchik AN, Grubov VV, Maksimenko VA, Lüttjohann A, Frolov NS, Marqués-Pascual C, et al. Extreme events in epileptic EEG of rodents after ischemic stroke. Eur Phys J Spec Top 2018;227(7-9):921-932. [doi: 10.1140/epjst/e2018-800019-1]
25. Frolov NS, Grubov VV, Maksimenko VA, Lüttjohann A, Makarov VV, Pavlov AN, et al. Statistical properties and predictability of extreme epileptic events. Sci Rep 2019 May 10;9(1):7243 [FREE Full text] [doi: 10.1038/s41598-019-43619-3] [Medline: 31076609]
26. Lhatoo S, Noebels J, Whittemore V, NINDS Center for SUDEP Research. Sudden unexpected death in epilepsy: identifying risk and preventing mortality. Epilepsia 2015 Nov;56(11):1700-1706 [FREE Full text] [doi: 10.1111/epi.13134] [Medline: 26494436]
27. Jobson JD. Applied Multivariate Data Analysis: Volume II: Categorical and Multivariate Methods. New York: Springer-Verlag; 2012.
28. Charbonnier S, Zoubek L, Lesecq S, Chapotot F. Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging. Comput Biol Med 2011 Jun;41(6):380-389. [doi: 10.1016/j.compbiomed.2011.04.001] [Medline: 21497802]
29. Hjorth B. EEG analysis based on time domain properties. Electroencephalogr Clin Neurophysiol 1970 Sep;29(3):306-310. [doi: 10.1016/0013-4694(70)90143-4] [Medline: 4195653]
30. Redmond S, Heneghan C. Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea. IEEE Trans Biomed Eng 2006 Mar;53(3):485-496. [doi: 10.1109/TBME.2005.869773] [Medline: 16532775]
31. Oh SH, Lee YR, Kim HN. A novel EEG feature extraction method using Hjorth parameter. Int J Electron Electr Eng 2014;2(2):106-110. [doi: 10.12720/ijeee.2.2.106-110]
32. Mallat S. A Wavelet Tour of Signal Processing. Amsterdam, Netherlands: Elsevier; 1999.
33. Misiti M, Misiti Y, Oppenheim G, Poggi JM. Luleå tekniska universitet, LTU. 1996. Wavelet Toolbox 4: User's Guide URL: https://www.ltu.se/cms_fs/1.51590!/wavelet%20toolbox%204%20user's%20guide%20(larger%20selection).pdf [accessed 2020-01-27]
34. Gysels E, Celka P. Phase synchronization for the recognition of mental tasks in a brain–computer interface. IEEE Trans Neural Syst Rehabil Eng 2004;12(4):406-415. [doi: 10.1109/tnsre.2004.838443]
35. Gouy-Pailler C, Achard S, Rivet B, Jutten C, Maby E, Souloumiac A, et al. Topographical Dynamics of Brain Connections for the Design of Asynchronous Brain-Computer Interfaces. In: Proceedings of the 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2007 Presented at: IEMBS'07; August 22-26, 2007; Lyon, France p. 2520-2523. [doi: 10.1109/iembs.2007.4352841]
36. Grosse-Wentrup M. Understanding Brain Connectivity Patterns during Motor Imagery for Brain-Computer Interfacing. In: Proceedings of the 2008 Conference on Neural Information Processing Systems. 2008 Presented at: NIPS'08; December 8-11, 2008; Vancouver, British Columbia, Canada URL: https://papers.nips.cc/paper/3505-understanding-brain-connectivity-patterns-during-motor-imagery-for-brain-computer-interfacing.pdf
37. Wei Q, Wang Y, Gao X, Gao S. Amplitude and phase coupling measures for feature extraction in an EEG-based brain-computer interface. J Neural Eng 2007 Jun;4(2):120-129. [doi: 10.1088/1741-2560/4/2/012] [Medline: 17409486]
38. Chung YG, Kim MK, Kim SP. Inter-channel Connectivity of Motor Imagery EEG Signals for a Noninvasive BCI Application. In: Proceedings of the 2011 International Workshop on Pattern Recognition in NeuroImaging.: IEEE; 2011 Presented at: PRNI'11; May 16-18, 2011; Seoul, South Korea. [doi: 10.1109/prni.2011.9]
39. Díaz MH, Córdova FM, Cañete L, Palominos F, Cifuentes F, Rivas G. Inter-channel correlation in the EEG activity during a cognitive problem solving task with an increasing difficulty questions progression. Procedia Comput Sci 2015;55:1420-1425. [doi: 10.1016/j.procs.2015.07.136]
40. Fraiwan L, Lweesy K, Khasawneh N, Wenz H, Dickhaus H. Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier. Comput Methods Programs Biomed 2012 Oct;108(1):10-19. [doi: 10.1016/j.cmpb.2011.11.005] [Medline: 22178068]
41. Li X, Cui L, Tao S, Chen J, Zhang X, Zhang G. HyCLASSS: a hybrid classifier for automatic sleep stage scoring. IEEE J Biomed Health Inform 2018 Mar;22(2):375-385. [doi: 10.1109/JBHI.2017.2668993] [Medline: 28222004]
42. Koley B, Dey D. An ensemble system for automatic sleep stage classification using single channel EEG signal. Comput Biol Med 2012 Dec;42(12):1186-1195. [doi: 10.1016/j.compbiomed.2012.09.012] [Medline: 23102750]
43. Breiman L. Random forests. Mach Learn 2001;45(1):5-32 [FREE Full text] [doi: 10.1023/A:1010933404324]
44. Längkvist M, Karlsson L, Loutfi A. Sleep Stage Classification Using Unsupervised Feature Learning. Adv Artif Neural Syst 2012;2012:1-9. [doi: 10.1155/2012/107046]
45. Rodríguez-Sotelo J, Osorio-Forero A, Jiménez-Rodríguez A, Cuesta-Frau D, Cirugeda-Roldán E, Peluffo D. Automatic sleep stages classification using EEG entropy features and unsupervised pattern analysis techniques. Entropy 2014;16(12):6573-6589. [doi: 10.3390/e16126573]
46. Makeig S, Bell AJ, Jung TP, Sejnowski TJ. Independent component analysis of electroencephalographic data. Adv Neural Inf Process Syst 1996;8(8):145-151 [FREE Full text]

47. Pavlov AN, Hramov AE, Koronovskii AA, Sitnikova EY, Makarov VA, Ovchinnikov AA. Wavelet analysis in neurodynamics. Phys-Usp 2012;55(9):845-875. [doi: 10.3367/ufne.0182.201209a.0905]
48. Gratton G, Coles MG, Donchin E. A new method for off-line removal of ocular artifact. Electroencephalogr Clin Neurophysiol 1983;55(4):468-484. [doi: 10.1016/0013-4694(83)90135-9]
49. Grubov VV, Runnova AE, Koronovskii AA, Hramov AE. Adaptive filtering of electroencephalogram signals using the empirical-modes method. Tech Phys Lett 2017;43(7):619-622. [doi: 10.1134/s1063785017070070]
50. Grubov VV, Sitnikova E, Pavlov AN, Koronovskii AA, Hramov AE. Recognizing of stereotypic patterns in epileptic EEG using empirical modes and wavelets. Physica A 2017 Nov;486:206-217. [doi: 10.1016/j.physa.2017.05.091]

Abbreviations

AE: amplitude energy
CSR: Center for Sudden Unexpected Death in Epilepsy Research
EEG: electroencephalogram
EMU: epilepsy monitoring unit
GTCS: generalized tonic-clonic seizures
ISW: intermittent slow wave
NIH: National Institutes of Health
PGES: postictal generalized electroencephalogram suppression
PPR_{5s}: 5-second tolerance-based positive prediction rate
SUDEP: sudden unexpected death in epilepsy

Edited by G Eysenbach, C Lovis; submitted 15.11.19; peer-reviewed by A Pisarchik, E Kutafina; comments to author 07.12.19; revised version received 20.12.19; accepted 29.12.19; published 14.02.20.

Please cite as:
Li X, Tao S, Jamal-Omidi S, Huang Y, Lhatoo SD, Zhang GQ, Cui L
Detection of Postictal Generalized Electroencephalogram Suppression: Random Forest Approach
JMIR Med Inform 2020;8(2):e17061
URL: https://medinform.jmir.org/2020/2/e17061
doi:10.2196/17061
PMID:32130173

Original Paper

The Impact of Electronic Health Records on the Duration of Patients' Visits: Time and Motion Study

Abdulrahman Mohammed Jabour[1], PhD

Department of Health Informatics, Faculty of Public Health and Tropical Medicine, Jazan University, Jazan, Saudi Arabia

Corresponding Author:
Abdulrahman Mohammed Jabour, PhD
Department of Health Informatics
Faculty of Public Health and Tropical Medicine
Jazan University
Almarifa Road
Jazan, 82822
Saudi Arabia
Phone: 966 17329500 ext 5545
Email: ajabour@jazanu.edu.sa

Abstract

Background: Despite the many benefits of electronic health records (EHRs), studies have reported that EHR implementation could create unintended changes in the workflow if not studied and designed properly. These changes may impact the time patients spend on the various steps of their visits, such as the time spent in the waiting area and with a physician. The amount of time patients spend in the waiting area before consultation is often a strong predictor of patient satisfaction, willingness to come back for a return visit, and overall experience. The majority of prior studies that examined the impact of EHR systems on time focused on single aspects of patient visits or user (physicians or nurses) activities. The impact of EHR use on patients' time spent during the different aspects of the visit is rarely investigated.

Objective: This study aimed to evaluate the impact of EHR systems on the amount of time spent by patients on different tasks during their visit to primary health care (PHC) centers.

Methods: A time and motion observational study was conducted at 4 PHC centers. The PHC centers were selected using stratified randomized sampling. Of the 4 PHC centers, 2 used an EHR system and 2 used a paper-based system. Each group had 1 center in a metropolitan area and another in a rural area. In addition, a longitudinal observation was conducted at one of the PHC centers after 1 year and again after 2 years of implementation. The analysis included descriptive statistics and group comparisons.

Results: The results showed no significant difference in the amount of time spent by patients in the reception area (P=.26), in the waiting area (P=.57), consultation time (P=.08), and at the pharmacy (P=.28) between the EHR and paper based groups. However, there was a significant difference (P<.001) in the amount of time spent on all tasks between the PHC centers located in metropolitan and rural areas. The longitudinal observation also showed reduction in the registration time (from 5.5 [SD 3.5] min to 0.9 [SD 0.5] min), which could be attributed to the introduction of a Web-based booking system.

Conclusions: The variation in the time patients spend at PHC centers is more likely to be attributed to the facility location than EHR use. The changes in the introduction of new tools and functions, however, such as the Web-based booking system, can impact the duration of patients' visits.

(JMIR Med Inform 2020;8(2):e16502) doi:10.2196/16502

KEYWORDS

patients experience; time and motion; waiting time; electronic health records

Introduction

Many studies have shown the benefits of electronic health records (EHRs) in reducing duplicate tests and procedures, reducing drug expenditures, improving the utilization of radiology tests, allowing for better documentation of charges, and decreasing billing errors [1-6]. On the other hand, some studies have reported that the new systems can disturb the current workflows and result in unintended consequences. The

Agency for Healthcare Research and Quality defined workflow as "a sequence of physical and mental tasks performed by various people within and between work environments. It can occur at several levels (one person, between people, across organizations) and can occur sequentially or simultaneously" [1].

The patients' waiting time and the consultation time are very important parts of patients' experience that could be impacted by the introduction of EHR systems. Many studies have shown that physicians are concerned about the amount of time needed for data entry, and the physicians have stated that the data entry time could be better used to provide direct patient care [2-7]. The distribution of patients' time during visits to primary health care (PHC) centers is a strong predictor of patient satisfaction and, thus, utilization. Studies have found that patients prefer to spend less time waiting for doctors, registering, or at the pharmacy and would prefer to have more time with physicians [8-10].

The vast majority of prior studies that investigated the impact of EHR on time can be categorized into two general classes of studies: efficiency studies and time and motion studies. Efficiency studies tend to focus on the number of patients who can be seen in a given period, whereas the majority of the EHR-related time and motion studies investigate the duration of a single task performed by health care providers [2-4,11-16]. Most patient-centered studies focused on the patient-physician interaction and the amount of time physicians allocate to patients. These studies examined the consultation time by comparing the time physicians allocate to EHR or electronic data entry with the amount of time physicians need for completing conventional paper-based documentation. Studies reported conflicting results regarding EHR's effects on consultation time [11,15,17]. In addition, one study also reported that more variation was attributed to the facility location than the system being implemented [11].

The results of these studies provided details about patients' experience and the amount of time spent at the doctor's office but did not provide information about the time spent before or after a physician visit. Examples of time spent before and after a physician visit include the time spent in the waiting room before seeing a physician. To determine the impact of EHR on patients, it is important to investigate the impact of EHR from a patient's perspective. The amount of time spent in a waiting area is strongly associated with patients' satisfaction and willingness to revisit [8-10,18-20]. Similarly, other tasks that do not involve interactions with a physician impact patients' satisfaction. These tasks could include registration and pharmacy services, which can add to the total duration of patients' visit.

The duration of users' experience with EHR can contribute to the duration of tasks at EHR-based facilities. Studies have indicated that user familiarity with a system is related to the amount of time per task. Some studies have highlighted reduced productivity in hospitals shortly after EHR implementation. The reduced productivity often improves as users become more familiar with the new system and develop the necessary skills to use the system efficiently. In some cases, the longer amount of time needed to perform tasks may continue, which can be explained by an EHR system having more functions and being more complicated than a comparable paper-based system [14]. The additional functions and features could result in a longer amount of time needed to complete tasks.

The aim of this study was to investigate the time patients spend at the various departments in PHC centers. The study focused on the following: time at registration, time spent in the waiting room, consultation time, and the time spent at the pharmacy. We hypothesized that the time patients spend at EHR-based and paper-based PHC centers is different. Furthermore, we hypothesized that the time patients spend at the EHR-based PHC centers will decrease with time after implementation.

Methods

Sites, Context, and Sampling

The Research Ethical Committee at Jazan University approved the study (approval number REC-39/4S005). We selected 4 PHC centers within Jazan area, Saudi Arabia, using a stratified randomized sampling. Of the 4 PHC centers, 2 used an EHR-based and 2 used a paper-based system. One of the 2 PHC centers using an her-based system was located in a metropolitan area and the other was located in a rural area. Similarly, 1 of the 2 PHC centers using a paper-based system was located in a metropolitan area and the other was located in a rural area.

Only public PHC centers operating under the Ministry of Health (MOH) were included in the study. These centers used the same policies and regulations related to funding, patient eligibility and coverage, and resources and were subject to the same laws. The MOH PHC centers provide public-free services to national citizens who make over 94% of the visitors, and the remaining noncitizens were covered through a private insurance or out-of-pocket [21]. Health care providers at the PHC centers included were general practitioners; some of the PHC centers provide basic dental services, which we excluded from this study.

As all PHC centers operate under the MOH, the 2 EHR-based PHC centers included were using the same EHR system, and the 2 paper-based PHC centers were using the same forms and documentation guidelines. Private and semipublic centers were excluded from the study to maintain homogeneity of sampling and to control for other confounding variables. More details about the PHC government solution strategy can be found on the Ministry of Health website [22].

Observation and Data Collection

The data collected included both cross-sectional and longitudinal observations. To investigate the impact of familiarity and facility experience with EHR on the time patients spend at each phase of the visit, longitudinal data were collected. The longitudinal observation was carried out at 1 of the EHR-based PHC centers, first at baseline during January 2018 and then at follow-up during December 2018. The remaining 3 PHC centers were observed cross-sectionally during the same period of the longitudinal follow-up observation (during November 2018 and December 2018).

The observation was conducted by 8 undergraduate health informatics interns. To ensure the consistency of data collection, observers were trained for 2 weeks on the workflow concepts, observation, and the data collection techniques of the study. Before data collection, they also spent 3 days at each site to familiarize themselves with the workflow processes. On the fourth day, they began the collection of data via direct observation. The duration of each task was documented using a stopwatch and papers. The observers were also sharing their findings and experiences with the research team on a weekly basis during the observation period.

The study focused on the time spent on each task from a patient's perspective and not the health provider's perspective. For example, a patient's waiting time is not a task that is based on the provider's activity time. Moreover, because the emphasis was on the time spent from a patient's perspective, the details of tasks or subtasks from a health care provider's perspective were not differentiated. For example, from a health care provider's perspective, patient registration involves the subtask of checking patient identities, entering patient information, and searching for the patient file. In this study, these tasks were considered as a single step—patient registration. In addition, if health care providers were performing parallel tasks, the duration of the tasks was measured as the time spent by the patient.

The definition of the beginning and end of each task was also defined from the patient's perspective. The reception time was defined as the time from the beginning of patient-clerk encounter to the end of the registration process. The waiting time was defined as the time from the end of the reception time to the time patients entered the physician's office. The consultation time was defined as the time from entering the physician's office to the time exiting the office. The pharmacy time was defined as the time from the beginning of patient-pharmacist encounter to the time patients receive the medication and instruction. An important point to highlight is that all waiting (before consultation) takes place in the waiting room. Some studies showed that a physician may serve multiple rooms sequentially by inviting patients to one of the rooms, getting their vital signs taken by the nurse, and then having them wait for the physician in the same room. We found that each physician has a single room and patients were instructed to enter the main physician's room when it was time to be seen.

For comparability, we also excluded the tasks that were not applicable to all PHC centers, such as the dental clinic or laboratory and x-ray. Patients who came for a dental visit or who end up going to the laboratory or x-ray department were excluded from the analysis. The reason for this is that patients who revisit the physician after an x-ray may have a different waiting time and consultation time from those visiting for an initial consultation.

Analysis

Data were analyzed using the Statistical Package for the Social Sciences version 21 (IBM Corp). The analysis performed included descriptive statistics to show the time per task for the different groups.

For the group comparison, we performed the Mann-Whitney test, a nonparametric test. The comparison included the time patients spent at each phase of their visit at the PHC centers. The factors evaluated were EHR system versus paper-based system and metropolitan area versus rural area.

The Mann-Whitney test was also used to examine the impact of PHC familiarity with the EHR system and the time patients spent on each task during visits. This included comparing the time per task during the baseline and follow-up.

Results

The results included comparing the time per task at the PHC centers using the EHR-based system with the PHC centers using the paper-based system and comparing the metropolitan PHC centers with the rural PHC centers. The results also included comparing the changes in the duration of tasks at 1 of the PHC centers after 1 year and after 2 years of using EHR.

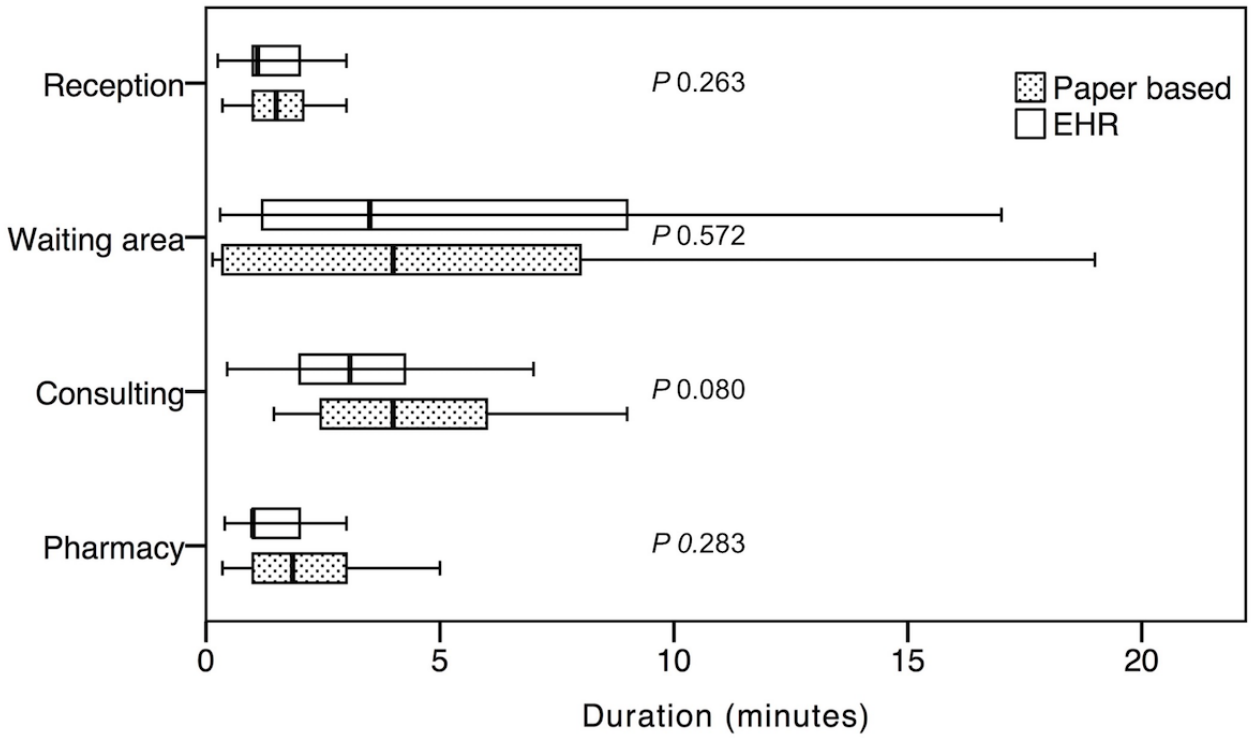
Electronic Health Record Versus Paper-Based Primary Health Care Centers

First, we observed the time spent performing basic tasks such as the registration time at the reception, time waiting for the physicians, consultation time, and time spent at the pharmacy. These observations were made at the 4 PHC centers. Other tasks such as getting x-rays and laboratory tests were excluded as these services were not available at all PHC centers. After the exclusion, the remaining number of events were 118 and 106 at the PHC centers using the EHR-based system and the PHC centers using the paper-based system, respectively. No significant differences were found between PHC centers that used an EHR-based system and those that used a paper-based system (P=.26, P=.57, P=.08, and P=.28 for the reception time, waiting time, consultation time, and time spent at the pharmacy, respectively; Table 1 and Figure 1).

Table 1. The time per task for primary health care centers using the electronic health record (EHR)–based system versus the paper-based system.

Task	EHR		Paper		*P* value
	Events, n	Time (min), mean (SD)	Events, n	Time (min), mean (SD)	
Reception	31	1.52 (1.02)	31	1.89 (1.36)	.26
Waiting for doctor	30	6.33 (7.37)	21	5.47 (6.11)	.57
Consultation time	28	3.30 (1.86)	26	6.39 (6.79)	.08
Pharmacy	29	1.61 (1.20)	28	1.95 (1.33)	.28

Figure 1. The time spent performing tasks at primary health care centers that use the electronic health record system and the paper-based system. EHR: electronic health record.



Rural Versus Metropolitan Primary Health Care Centers

The time spent performing tasks at each of the 4 PHC centers was also compared based on the location of the PHC centers (Figure 2). There were 2 metropolitan PHC centers and 2 rural PHC centers with 109 and 115 events, respectively. Our results showed statistically significant difference between the PHC centers located in metropolitan areas and the PHC centers located in rural areas for all 4 tasks, with $P<.001$ (Table 2).

Figure 2. The time spent performing tasks at primary health care centers in metropolitan and rural areas.

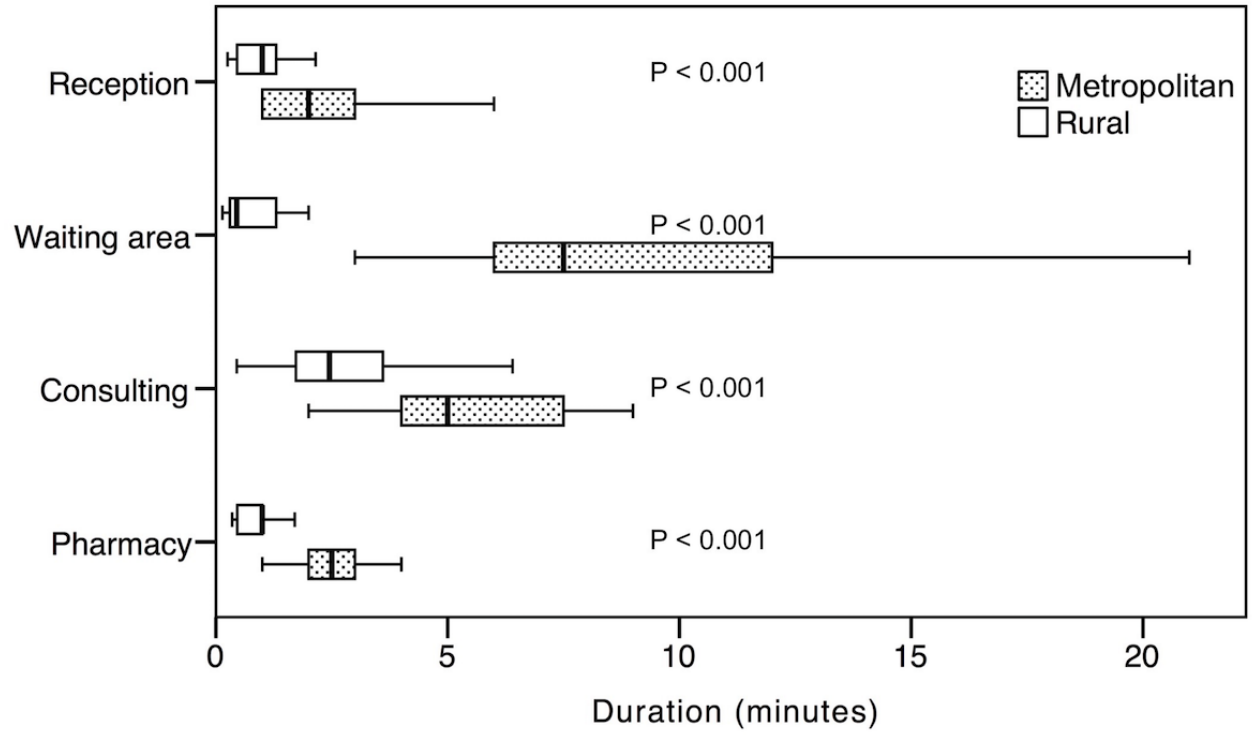


Table 2. The time per task for primary health care centers located in metropolitan areas versus primary health care centers located in rural areas.

Task	Metropolitan		Rural		*P* value
	Number of events, n	Time (min), mean (SD)	Number of events, n	Time (min), mean (SD)	
Reception	30	2.36 (1.299)	32	1.09(0.710)	<.001
Waiting for doctor	26	9.73 (6.20)	25	2.07 (5.10)	<.001
Consultation time	23	7.43 (6.79)	31	2.83 (1.64)	<.001
Pharmacy	30	2.60 (1.22)	27	0.86 (0.42)	<.001

Longitudinal Observation for Electronic Health Record–Based Primary Health Care Center

Finally, we examined the effect of familiarity with the EHR on the duration of tasks. At one of the EHR-based PHC centers, the observation was conducted longitudinally (Figure 3). There were 72 events at the baseline observation and 63 events at the 12-month follow-up. When comparing the time patients spent at each phase during their visits at baseline and follow-up, there was a significant difference in the time spent at the reception (P<.001) and at the pharmacy (P=.01). The difference in time patients spent waiting for the doctor and the consultation time was insignificant, with P=.22 and P=.36, respectively (Table 3).

Figure 3. The time spent performing tasks at 1 electronic health record–based primary health care center after 1 year.

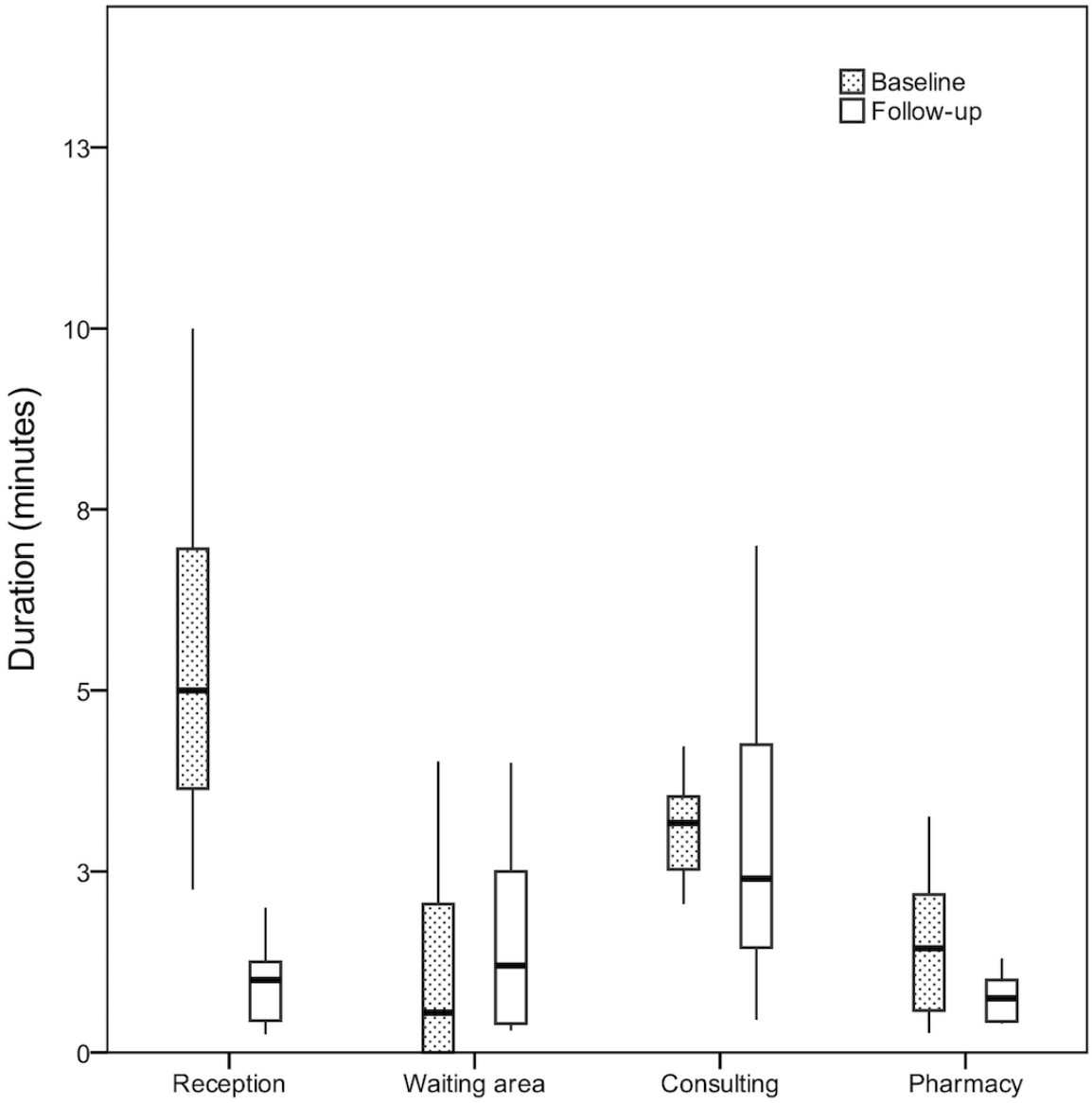


Table 3. The time spent per task at baseline and 12-month follow-up at 1 electronic health record–based primary health care center.

Task	Before		After		*P* value
	Number of events, n	Time (min), mean (SD)	Number of events, n	Time (min), mean (SD)	
Reception	15	5.5 (3.5)	16	0.9 (0.5)	<.001
Waiting for doctor	15	1.87 (3.5)	15	3.79 (7.89)	.22
Consultation time	20	3.1 (0.8)	18	3.05 (2.08)	.36
Pharmacy	22	1.49 (0.94)	14	0.75 (0.33)	.01

Discussion

Overview

The vast majority of prior research examined the impact of EHR on the duration of tasks from the user's perspective (such as examining the time taken by doctors and other hospital staff to perform tasks). They often focus on a particular task or subtask that was aided by a new tool such as the computerized provider order entry or personal digital assistant. The impact of these tools on a patient's time was rarely examined from the patient's perspective [14]. In this study, we examined the difference in time taken per task for patients at PHC centers that used an EHR-based system compared with those that used a paper-based system. We also compared the time taken per task at PHC centers in the metropolitan and rural areas. Then, we examined the differences in task duration in 1 year and 2 years of EHR implementation.

One advantage of our study is that it controlled for many of the confounding variables that could vary based on the PHC centers. Some of these variables were the facility rules and regulations, funding, the type of system used (either EHR or paper form), and patient eligibility. We also applied the same data collection techniques using time and motion observations for both groups instead of using artifacts or a timestamp analysis.

Compared with prior studies, the average waiting time was relatively short. In a study conducted in the United States for instance, the average waiting time for a family physician was 13.5 min [23]. In our study, the average waiting time was 6.33, 5.47, 9.73, and 2.07 min for the PHC centers using an EHR-based system, using a paper-based system, in the metropolitan areas, and in the rural areas, respectively. Consistent with our result, a study conducted in Saudi reported that 83% of patients had a waiting time of less than 5 min [24].

Consistent with our findings, prior studies showed variations in the waiting time based on the geographical area [25]. Studies also reported that the distribution of PHC centers in the country was consistent with population distribution. This distribution resulted in overutilization of some PHC centers and underutilization of others [26]. The overutilization or underutilization of certain PHC centers could help explain why certain patients were unsatisfied with the waiting time at particular locations [8,20]. The concept of *satisfaction* was based on a self-administered survey, and the survey did not inquire about the exact waiting time and also did not provide documentation of the waiting time based on direct observations or EHR audit files [20].

Consultation time was an important factor as it impacts the quality of care, patient satisfaction, and level of utilization [18,27-29]. Although consultation time varies based on the country, studies showed that patients, in general, prefer a longer consultation time [18]. Studies based in the United States reported an average consultation time of 10 to 15 min, whereas a local study reported that 80% to 85% of patients spent less than 5 min with the doctor and 10% to 16% of patients spent 5 to 10 min with the doctor [24].

Comparing consultation times in this study with consultation times in prior studies must be done cautiously, as the local studies are outdated, and those that were conducted in the United States follow different workflow practices. In the United States, patients typically see a nurse who will take basic vital signs, collect medical history, and obtain general signs and symptoms [25]. Following this interaction with a nurse, patients will then wait at the doctor's office to be seen [25]. This waiting time is sometimes counted as part of the consultation time, which will result in a longer patient-doctor interaction. In the PHC centers where our study was conducted, patients are called directly into the room to see the doctor, and the visit with the doctor begins at this point with no waiting time in between [25].

Consistent with prior studies, no significant difference was found in the duration of tasks between the PHC centers that use EHR-based systems and the PHC centers that use paper-based systems [12,17]. A significant difference was found in the duration of tasks between PHC centers based on location.

Our results did not show any significant difference in patients' waiting time or consultation time after 2 years of EHR adoption. There was a significant difference in the time spent registering and at the pharmacy. The time spent at the reception decreased from an average of 5.5 min (SD 3.5) in January 2018 to an average of 0.9 min (SD 0.5) in December 2018. The decrease in time could be attributed to the MOH Web-based booking system, which was adopted between January 2018 and December 2018. The MOH Web-based booking system called Mawid was implemented to allow patients to book, cancel, or reschedule appointments while also allowing individuals to manage referral appointments [30]. This booking service was provided by the government as part of a larger initiative, which was intended to help to verify patient identities by linking patients to a national ID. The Web-based booking system required patients to enter the information needed by the registration office online before visiting the PHC center. Although this service was provided initially as an optional service, many PHC centers have made the service mandatory for accepting nonurgent patients.

Limitations and Future Work

Although we tried to control for confounding variables such as the type of EHR system being used, facility rules and regulations, funding, and patient eligibility, we did not account for some factors that could impact the generalizability of our study. One of the factors was patients' conditions and demographics. Patients with more complex diseases may require more time for consultation and data entry. In addition, the type of EHR system being used can impact the outcome. Although the system being used at PHC centers is provided and approved by MOH [22], the generalizability of our finding to PHC centers that use a different EHR system is unknown. Moreover, we were unable to determine the effect of system familiarity on the reception time for the longitudinal part of the study because of the introduction of the Web-based booking system after the baseline period. Another limitation is that we did not measure the interobserver reliability. However, all observers were similar in their academic qualification, experience, and training received before data collection. For future studies, we recommend controlling for patients' conditions and reasons for visits because of their expected impact on the duration of visits. This will not only help in explaining the source of variation in time but also improve the generalizability of the results.

Our result shows a significant difference in the duration of tasks between metropolitan and rural PHC centers; however, the cause of these differences is yet unknown. This could be further investigated in future studies by including more PHC centers and examining the potential factors such as reason for visits, staffing, and variation in workflow and clinical practice. Moreover, we recommend examining the waiting time and consultation time in the different cities within the country. In addition, more studies are needed to examine the impact of EHR on the way patients spend their time when visiting the doctor in more busy environments.

Conclusions

Our study showed that the time spent by patients on the various tasks during PHC center visits is the same at both EHR- and paper-based PHC centers. We also found that patients' waiting time and consultation time were the same after 1 year and 2 years of EHR implementation. The registration time, however, decreased when comparing the time after 1 year with the time after 2 years of EHR implementation. We expect that the change was attributed to the Web-based booking systems rather than EHR itself. Apart from the training and skills related to short-term impact after EHR implementation, we believe that changes in time after EHR use are often attributed to the addition or elimination of tasks and functions rather than EHR itself. Therefore, focusing on the EHR function that minimizes the tasks performed by patients can shorten the duration of their visits and enhance their satisfaction. Some of these tasks include Web-based tools for booking, entry of patients' history, and medication refill.

Acknowledgments

The author would like to thank the research assistants for their time and effort during the data collection. The author also wants to acknowledge Jazan University for their funding support. This research was supported by the Deanship of Scientific Research at Jazan University. The deanship has no role in the design, analysis, or interpretation of the results.

Conflicts of Interest

None declared.

References

1. Agency for Healthcare Research and Quality. AHRQ Health IT. 2019. Workflow Assessment for Health IT Toolkit URL: https://healthit.ahrq.gov/health-it-tools-and-resources/evaluation-resources/workflow-assessment-health-it-toolkit/workflow [accessed 2019-12-16]
2. Hsu J, Huang J, Fung V, Robertson N, Jimison H, Frankel R. Health information technology and physician-patient interactions: impact of computers on communication during outpatient primary care visits. J Am Med Inform Assoc 2005;12(4):474-480 [FREE Full text] [doi: 10.1197/jamia.M1741] [Medline: 15802484]
3. Alkureishi MA, Lee WW, Lyons M, Press VG, Imam S, Nkansah-Amankra A, et al. Impact of electronic medical record use on the patient-doctor relationship and communication: a systematic review. J Gen Intern Med 2016 May;31(5):548-560 [FREE Full text] [doi: 10.1007/s11606-015-3582-1] [Medline: 26786877]
4. Frankel R, Altschuler A, George S, Kinsman J, Jimison H, Robertson NR, et al. Effects of exam-room computing on clinician-patient communication: a longitudinal qualitative study. J Gen Intern Med 2005 Aug;20(8):677-682 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0163.x] [Medline: 16050873]
5. Joukes E, de Keizer N, Abu-Hanna A, de Bruijne M, Cornet R. End-user experiences and expectations regarding data registration and reuse before the implementation of a (new) electronic health record: a case study in two university hospitals. Stud Health Technol Inform 2015;216:997. [Medline: 26262299]
6. Lee WW, Alkureishi MA, Ukabiala O, Venable LR, Ngooi SS, Staisiunas DD, et al. Patient perceptions of electronic medical record use by faculty and resident physicians: a mixed methods study. J Gen Intern Med 2016 Nov;31(11):1315-1322 [FREE Full text] [doi: 10.1007/s11606-016-3774-3] [Medline: 27400921]
7. Rathert C, Mittler JN, Banerjee S, McDaniel J. Patient-centered communication in the era of electronic health records: what does the evidence say? Patient Educ Couns 2017 Jan;100(1):50-64. [doi: 10.1016/j.pec.2016.07.031] [Medline: 27477917]

8. Bielen F, Demoulin N. Waiting time influence on the satisfaction - loyalty relationship in services. Manag Serv Qual Int J 2007;17(2):174-193. [doi: 10.1108/09604520710735182]
9. Camacho F, Anderson R, Safrit A, Jones AS, Hoffmann P. The relationship between patient's perceived waiting time and office-based practice satisfaction. N C Med J 2006;67(6):409-413. [Medline: 17393701]
10. Xie Z, Or C. Associations between waiting times, service times, and patient satisfaction in an endocrinology outpatient department: A time study and questionnaire survey. Inquiry 2017;54:46958017739527 [FREE Full text] [doi: 10.1177/0046958017739527] [Medline: 29161947]
11. Joukes E, Abu-Hanna A, Cornet R, de Keizer N. Time spent on dedicated patient care and documentation tasks before and after the introduction of a structured and standardized electronic health record. Appl Clin Inform 2018 Jan;9(1):46-53 [FREE Full text] [doi: 10.1055/s-0037-1615747] [Medline: 29342479]
12. Overhage JM, Perkins S, Tierney WM, McDonald CJ. Controlled trial of direct physician order entry: effects on physicians' time utilization in ambulatory primary care internal medicine practices. J Am Med Inform Assoc 2001;8(4):361-371 [FREE Full text] [doi: 10.1136/jamia.2001.0080361] [Medline: 11418543]
13. Lo HG, Newmark LP, Yoon C, Volk LA, Carlson VL, Kittler AF, et al. Electronic health records in specialty care: a time-motion study. J Am Med Inform Assoc 2007;14(5):609-615 [FREE Full text] [doi: 10.1197/jamia.M2318] [Medline: 17600102]
14. Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health records on time efficiency of physicians and nurses: a systematic review. J Am Med Inform Assoc 2005;12(5):505-516. [doi: 10.1197/jamia.M1700] [Medline: 15905487]
15. Scott PJ, Curley PJ, Williams PB, Linehan IP, Shaha SH. Measuring the operational impact of digitized hospital records: a mixed methods study. BMC Med Inform Decis Mak 2016 Nov 10;16(1):143 [FREE Full text] [doi: 10.1186/s12911-016-0380-6] [Medline: 27829453]
16. Holman GT, Beasley JW, Karsh BT, Stone JA, Smith PD, Wetterneck TB. The myth of standardized workflow in primary care. J Am Med Inform Assoc 2016 Jan;23(1):29-37 [FREE Full text] [doi: 10.1093/jamia/ocv107] [Medline: 26335987]
17. Pizziferri L, Kittler A, Volk L, Honour MM, Gupta S, Wang S, et al. Primary care physician time utilization before and after implementation of an electronic health record: a time-motion study. J Biomed Inform 2005 Jun;38(3):176-188 [FREE Full text] [doi: 10.1016/j.jbi.2004.11.009] [Medline: 15896691]
18. Deveugele M, Derese A, van den Brink-Muinen A, Bensing J, de Maeseneer J. Consultation length in general practice: cross sectional study in six European countries. Br Med J 2002 Aug 31;325(7362):472 [FREE Full text] [doi: 10.1136/bmj.325.7362.472] [Medline: 12202329]
19. Mohamed EY, Sami W, Alotaibi A, Alfarag A, Almutairi A, Alanzi F. Patients' satisfaction with primary health care centers' services, Majmaah, kingdom of Saudi of Saudi Arabia. Int J Health Sci (Qassim) 2015 Apr;9(2):163-170 [FREE Full text] [doi: 10.12816/0024113] [Medline: 26309435]
20. Mansour AA, al-Osimy MH. A study of satisfaction among primary health care patients in Saudi Arabia. J Community Health 1993 Jun;18(3):163-173. [doi: 10.1007/bf01325160] [Medline: 8408747]
21. Ministry of Health. The Saudi Ministry of Health. Visitors of Health Centers, MOH by Region, Clinic, % of Saudi and % of Cases Referred to Hospitals -2009 URL: https://data.gov.sa/Data/en/dataset/of-health-centers--moh-by-region--clinic----of-saudi-and---of-cases-referred-to-hospitals--2009 [accessed 2019-12-16]
22. Ministry of Health. The Saudi Ministry of Health. 2019. The New PHC Systems: National E- Health Strategy URL: https://www.moh.gov.sa/en/Ministry/nehs/Pages/The-New-PHC-Systems.aspx [accessed 2019-12-16]
23. Hirsch AG, Jones JB, Lerch VR, Tang X, Berger A, Clark DN, et al. The electronic health record audit file: the patient is waiting. J Am Med Inform Assoc 2017 Apr 1;24(e1):e28-e34. [doi: 10.1093/jamia/ocw088] [Medline: 27375293]
24. al-Faris EA, al-Dayel MA, Ashton C. The effect of patients' attendance rate on the consultation in a health centre in Saudi Arabia. Fam Pract 1994 Dec;11(4):446-452. [doi: 10.1093/fampra/11.4.446] [Medline: 7895975]
25. Oostrom T, Einav L, Finkelstein A. Outpatient office wait times and quality of care for medicaid patients. Health Aff (Millwood) 2017 May 1;36(5):826-832 [FREE Full text] [doi: 10.1377/hlthaff.2016.1478] [Medline: 28461348]
26. Al-Jaber A, Da'ar OB. Primary health care centers, extent of challenges and demand for oral health care in Riyadh, Saudi Arabia. BMC Health Serv Res 2016 Nov 4;16(1):628 [FREE Full text] [doi: 10.1186/s12913-016-1876-6] [Medline: 27809919]
27. Cape J. Consultation length, patient-estimated consultation length, and satisfaction with the consultation. Br J Gen Pract 2002 Dec;52(485):1004-1006 [FREE Full text] [Medline: 12528588]
28. Ahmad BA, Khairatul K, Farnaza A. An assessment of patient waiting and consultation time in a primary healthcare clinic. Malays Fam Physician 2017;12(1):14-21 [FREE Full text] [Medline: 28503269]
29. Pillay DI, Ghazali RJ, Manaf NH, Abdullah AH, Bakar AA, Salikin F, et al. Hospital waiting time: the forgotten premise of healthcare service delivery? Int J Health Care Qual Assur 2011;24(7):506-522. [doi: 10.1108/09526861111160553] [Medline: 22204085]
30. Ministry of Health. The Saudi Ministry of Health. E-Services: (Mawid) Service URL: https://www.moh.gov.sa/en/eServices/Pages/cassystem.aspx [accessed 2019-12-16]

Abbreviations

EHR: electronic health record
MOH: Ministry of Health
PHC: primary health care

Edited by C Lovis; submitted 04.10.19; peer-reviewed by A Ekeland, L Rusu; comments to author 10.11.19; revised version received 29.11.19; accepted 01.12.19; published 07.02.20.

Please cite as:
Jabour AM
The Impact of Electronic Health Records on the Duration of Patients' Visits: Time and Motion Study
JMIR Med Inform 2020;8(2):e16502
URL: http://medinform.jmir.org/2020/2/e16502/
doi:10.2196/16502
PMID:

Original Paper

Optimizing Antihypertensive Medication Classification in Electronic Health Record-Based Data: Classification System Development and Methodological Comparison

Caitrin W McDonough[1], MS, PhD; Steven M Smith[1], PharmD, MPH; Rhonda M Cooper-DeHoff[1,2], PharmD, MS; William R Hogan[3], MD, MS

[1]Department of Pharmacotherapy and Translational Research, College of Pharmacy, University of Florida, Gainesville, FL, United States
[2]Division of Cardiovascular Medicine, Department of Medicine, College of Medicine, University of Florida, Gainesville, FL, United States
[3]Department of Health Outcomes & Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, United States

Corresponding Author:
Caitrin W McDonough, MS, PhD
Department of Pharmacotherapy and Translational Research
College of Pharmacy
University of Florida
PO Box 100486
Gainesville, FL,
United States
Phone: 1 3522736435
Fax: 1 3522736121
Email: cmcdonough@cop.ufl.edu

Abstract

Background: Computable phenotypes have the ability to utilize data within the electronic health record (EHR) to identify patients with certain characteristics. Many computable phenotypes rely on multiple types of data within the EHR including prescription drug information. Hypertension (HTN)-related computable phenotypes are particularly dependent on the correct classification of antihypertensive prescription drug information, as well as corresponding diagnoses and blood pressure information.

Objective: This study aimed to create an antihypertensive drug classification system to be utilized with EHR-based data as part of HTN-related computable phenotypes.

Methods: We compared 4 different antihypertensive drug classification systems based off of 4 different methodologies and terminologies, including 3 RxNorm Concept Unique Identifier (RxCUI)–based classifications and 1 medication name–based classification. The RxCUI-based classifications utilized data from (1) the Drug Ontology, (2) the new Medication Reference Terminology, and (3) the Anatomical Therapeutic Chemical Classification System and DrugBank, whereas the medication name–based classification relied on antihypertensive drug names. Each classification system was applied to EHR-based prescription drug data from hypertensive patients in the OneFlorida Data Trust.

Results: There were 13,627 unique RxCUIs and 8025 unique medication names from the 13,879,046 prescriptions. We observed a broad overlap between the 4 methods, with 84.1% (691/822) to 95.3% (695/729) of terms overlapping pairwise between the different classification methods. Key differences arose from drug products with multiple dosage forms, drug products with an indication of benign prostatic hyperplasia, drug products that contain more than 1 ingredient (combination products), and terms within the classification systems corresponding to retired or obsolete RxCUIs.

Conclusions: In total, 2 antihypertensive drug classifications were constructed, one based on RxCUIs and one based on medication name, that can be used in future computable phenotypes that require antihypertensive drug classifications.

(JMIR Med Inform 2020;8(2):e14777) doi:10.2196/14777

KEYWORDS

antihypertensive agents; electronic health records; classification; RxNorm; phenotype

Introduction

Background

Electronic health records (EHRs) contain a wealth of clinical information, and the development of tools to perform structured queries of common data models (CDMs) to standardized data formats has allowed researchers to utilize EHR data for discriminating complex clinical phenotypes with high measurement validity [1-4]. Measuring such complex phenotypes often requires integration of multiple streams of EHR data, including diagnoses or procedures, clinical measurements (eg, vitals and laboratory parameters), and medications or other exposures [5]. Additionally, the accurate measurement of these complex phenotypes relies on accurate measurement and classification of these individual components [6,7]. The classification of medication use in this context offers unique challenges because many medications have multiple indications and dosage forms. Furthermore, existing therapeutic classification systems generally group drug entities in ways that may only partially correlate with intended use [8], eg, numerous medications block beta-adrenergic receptors (beta-blockers), and such an effect can lower blood pressure (BP), but not all beta-blockers are used clinically as antihypertensives. Even among beta-blockers that are used clinically as antihypertensives, not all dosage forms effectively treat systemic hypertension (HTN) [9]. Thus, the mere presence of a drug entity in the prescribing record may not be sufficient to meet medication-related criteria in a complex phenotype.

Objectives

Accordingly, we aimed to design a set of standardized antihypertensive drug codes to be usable and maintainable by both ourselves and others in the research community [10]. We performed this study in the context of optimizing medication classification for use in HTN-related computable phenotypes. To properly identify patients who are taking antihypertensive medications, the prescription drug information from the EHR must be properly classified into antihypertensive therapeutic indication and antihypertensive drug classes. We designed our study by utilizing concepts (RxNorm and medication name) present in the prescription drug information from data models common in the United States (Informatics for Integrating Biology & the Bedside [i2b2], The National Patient Centered Clinical Research Network [PCORnet] CDM, and The Observational Medical Outcomes Partnership [OMOP] CDM) [1,3,4]. The main objective of this study was to report the development of a set of standardized drug codes and names for use in querying EHR data for antihypertensive medication prescriptions. The second objective was to compare different methods and resources for creating the antihypertensive drug classification. The third and final objective was to explore the coverage of different methods and resources (for creating the classification) on querying a large EHR-based dataset.

Methods

Resources

RxNorm

RxNorm is a standardized terminology to represent drugs. It was developed by the US National Library of Medicine (NLM) in 2002 and represents medications through normalized names for clinical drugs, which include ingredient or ingredients strength or strengths, and dose form [11,12]. The normalized names allow equivalent drug terms from different source vocabularies to be grouped together under the same RxNorm Concept Unique Identifier (RxCUI) [11]. RxNorm was designed to support electronic prescribing, mapping between different drug vocabularies, and the development of medication-related clinical decision support rules [11,12]. RxNorm can be accessed through an application programming interface (API), the RxNorm Navigator (RxNav) [12-14]. Currently, RxNorm integrates drug terminology from many sources including most drug knowledge bases (eg, Multum, Micromedex, and First DataBank), standard terminologies (eg, SNOMED CT and MeSH), the new Medication Reference Terminology (MED-RT), federal agencies in the United States (eg, Food and Drug Administration Structured Product labels), and international drug resources such as the Anatomical Therapeutic Chemical (ATC) Classification System and DrugBank [11,15].

Drug Ontology

The Drug Ontology (DrOn) is a formal representation of drug products, drug ingredients, mechanisms of action, therapeutic indications, strengths, and dosage forms based on the OWL2 Web Ontology Language [8,16-18]. These representations were created to allow researchers to query drug datasets, which usually come from EHRs or health insurance claim databases and typically use RxCUIs to identify different aspects of drug products (ie, ingredient or ingredients, dose forms, and strength or strengths), and National Drug Codes to identify individual drug products and the manufacturer and packaging thereof [16,17]. Currently, DrOn represents a drug's therapeutic indication as being a property of a drug product, which is a composite of one or more drug ingredients and excipients, whereas its mechanism of action (MoA) is represented as a property of the chemical compound or chemical compounds that constitutes the drug's ingredients [8,16,17]. DrOn has a hand-curated component and a component built automatically from RxNorm [16-18]. At present, DrOn contains representations for several mechanisms of action belonging to various antihypertensive drugs classes [8,16,17]. It also contains class representations for the drug products and ingredients, dosage forms, strengths, and therapeutic indications.

Data Source

OneFlorida and the OneFlorida Hypertension Population

The OneFlorida Clinical Research Consortium is a statewide network of health systems, providers, and payers covering more than 74% of Florida's population [19]. The network's catchment area covers all 67 Florida counties and allows for the facilitation

of clinical and translational research in health care settings and communities throughout the state. OneFlorida houses a Data Trust, which contains longitudinal EHR data on approximately 14 million Floridians (approximately 66%), mapped to the PCORnet CDM [19,20]. The HTN population within OneFlorida was defined as all adults (aged ≥18 years) with ≥1 HTN diagnosis from an outpatient encounter, defined as International Classification of Diseases (ICD)-9 code 401.x (Essential HTN) or ICD-10 code I10 (Essential [primary] HTN). The data utilized for this study were extracted from the OneFlorida Data Trust on December 14, 2017, in the PCORnet CDM, version 3.0, and included EHR data from June 2000 to July 2017, with 99.97% from encounters occurring from January 2011 onward. All (100%) of the prescription drug data were from January 2011 onward.

Prescription Drug Data

All prescription drug data for the HTN population were extracted from the Prescribing Table, including information on raw medication name and RxCUI. A total of 2 drug lists were created from this dataset. The first contained all unique raw medication names (see Drug Classification by Ingredient Name) derived directly from source EHRs, ie, not cleaned or curated during mapping to the PCORnet CDM. The second contained all unique RxCUIs (see Drug Classification by RxCUI utilizing DrOn and Drug Classification by RxCUI utilizing RxClass), which may be derived from source EHRs or created during mapping to the PCORnet CDM.

Drug Classification by Ingredient Name

A Medication Name Classification was constructed for antihypertensive medications, utilizing drug ingredient names. A summary of the features included is available in Multimedia Appendix 1. Both brand name and generic name were included. The list was constructed through manual curation, including methods from prior antihypertensive drug classifications [21,22]. The manually curated list was also reviewed by authors with biomedical informatics expertise (CM) and HTN pharmacotherapy expertise (CM and RC). The Medication Name Classification contains 286 drug ingredient names, further classified by antihypertensive drug class (eg, beta-blockers, angiotensin II receptor inhibitors [ARBs], calcium channel blockers [CCBs], etc).

To apply the Medication Name Classification to the OneFlorida raw medication list, the first word in the field was extracted as the ingredient name, with additional coding to capture combination medications (eg, adding underscores between the first and second words) and strings where the ingredient was not the first word. The antihypertensive Medication Name Classification was merged with the unique raw medication names from the OneFlorida dataset to map the drugs by antihypertensive drug class. All of the raw medication names that did not merge with the Medication Name Classification were discarded (eg, statins, insulin, etc).

Drug Classification by RxNorm Concept Unique Identifier Utilizing the Drug Ontology

The DrOn RxCUI Classification was constructed through multiple steps (Multimedia Appendix 1). First, all RxCUIs with a therapeutic indication for HTN (antihypertensive function) were extracted from DrOn using the *dron-query* tool [8]. Then, separate lists of RxCUIs were extracted for drugs with any of the following mechanisms of action: angiotensin-converting enzyme (ACE)–inhibitor, ARB, beta-blocker, CCB, loop diuretic, and thiazide and thiazide-like diuretic. These lists were then merged to assign an antihypertensive MoA to each RxCUI with a therapeutic indication for HTN. Combination drug products were assigned multiple mechanisms of action, representing each ingredient. The list was then manually reviewed, and mechanisms of action were added for drug products with mechanisms of action not currently represented in DrOn. These included the following: aldosterone antagonists, direct renin inhibitors, alpha-1 blockers, potassium-sparing diuretics, vasodilators, centrally acting agents, and other agents. The list was then reviewed by authors with biomedical informatics expertise (CM and WH) and HTN pharmacotherapy expertise (CM, SS, and RC). The DrOn RxCUI Classification contains 2543 antihypertensive RxCUIs, of SCDF, SCD, and SBD term types, organized by antihypertensive drug class or drug classes.

The unique RxCUIs from the OneFlorida dataset were merged with the DrOn RxCUI Classification to map the drugs by antihypertensive drug class. All of the RxCUIs that did not merge with the DrOn RxCUI Classification were discarded (eg, statins, insulin, etc).

Drug Classification by RxNorm Concept Unique Identifier Utilizing RxClass

The unique list of RxCUIs extracted from OneFlorida were also mapped utilizing the RxClass API on RxMix [23]. RxCUIs with less than 4 digits were removed (n=25). The function, “getClassByRxNormDrugId,” was used to obtain the drug classes for a specified drug identifier. In total, 2 different relationship sources (relaSource) were tested: ATC and MED-RT (Multimedia Appendix 1). Within MED-RT, the following relationships (rela) were selected: “has_MoA” and “may_treat.” The unique RxCUIs from the OneFlorida dataset were classified using the ATC and MED-RT relationship sources through the batch input mode.

Comparison Between Drug Classifications

The different drug classification methods were compared pairwise by calculating percent coverage and by reviewing the overlapping and nonoverlapping sets of RxCUIs among them. For all classification methods, the percent of antihypertensive drugs covered was calculated as the number of antihypertensive medications mapped by the classification method divided by the total number of unique terms (Raw Name or RxCUI) contained in the OneFlorida Prescribing Table. Within the ATC relationship source from RxClass, the antihypertensive classes were selected from the “name” field. A complete list of the ATC relationships included as antihypertensive drugs is available in Multimedia Appendix 2. Within the MED-RT relationship source from RxClass, 2 steps were used to select antihypertensive drugs: (1) RxCUIs were selected with the “may_treat” HTN relationship and (2) RxCUIs were further filtered based on the “has_MoA” relationship, including only those drugs from antihypertensive classes. A complete list of

the MED-RT "has_MoA" relationships included as antihypertensive drugs is available in Multimedia Appendix 3. Within the MED-RT relationship source from RxClass, the results from "may_treat" relationship were used to calculate total coverage. Differences between the antihypertensive coverage results were identified by pairwise comparisons and merges between the DrOn RxCUI Classification and every other classification.

On the basis of our review of the overlapping and nonoverlapping sets of names and RxCUIs among the classifications, we created a version 1.0 of 2 finalized classifications—one for ingredient names and one for RxCUIs—for use by us and other researchers. We published them on GitHub [10] to make them findable and reusable and to enable community contributions to future versions.

Application of the Drug Classifications to OneFlorida

The Medication Name Classification and the DrOn RxCUI Classification were applied to all of the prescription drug data from the OneFlorida HTN population. All available prescriptions from January 2011 onward were considered. Antihypertensive coverage by each method was calculated as number of prescription records mapped to an antihypertensive drug class by each map divided by the total number of prescriptions. Differences between the coverage results were identified by pairwise comparison. Additionally, summary-level frequency counts and percentages by the antihypertensive drug class were also calculated. All coding for mapping the drug data and the summary statistics was conducted using SAS version 9.4 (SAS).

Results

Data Source

At the time of the data extraction, there were 1,188,977 patients in the OneFlorida Data Trust with an ICD-9 or ICD-10 diagnosis code for HTN (Table 1).

In total, there were 13,879,046 prescriptions from these patients over the study period (January 2011 to July 2017; approximately 6.6 years). These prescriptions consisted of 13,627 unique RxCUIs and 8025 unique first words from the raw medication name string (Table 1).

Table 1. Counts of data attributes in the OneFlorida hypertensive patient prescribing table dataset.

Data attributes	Values, N
Hypertensive patients	1,188,977
Prescription records	13,879,046
Unique RxCUIs[a]	13,627
Unique Raw Med Name[b]	8025

[a]RxCUI: RxNorm Concept Unique Identifier.

[b]Unique first word of the Raw_Med_Name field, after additional data cleaning and quality control steps.

Drug Classification by Ingredient Name

The initial Medication Name Classification contained 286 antihypertensive medications that are mapped to 35 antihypertensive medication classes or combination medication classes (eg, ACE inhibitors, ARBs, CCBs, CCB-ARB combinations, etc). We chose to exclude timolol to be conservative. On the basis of this classification system, it is impossible to distinguish between oral and ophthalmic products. An excerpt from the Medication Name Classification is shown in Table 2, whereas the full map is available in Multimedia Appendix 4. Each entry consists of an arbitrary code, the single word drug name, the generic name (if applicable), and the drug class.

Table 2. Excerpt from Medication Name Classification.

Code[a]	Drug_Name	Generic[b]	Drug_Class
1	Benazepril	—[c]	ACE[d]
2	Lotensin	benazepril	ACE
3	Captopril	—	ACE
4	Capoten	captopril	ACE
5	Enalapril	—	ACE
6	Enalaprilat	enalapril	ACE
7	Fosinopril	—	ACE
8	Monopril	fosinopril	ACE
9	Lisinopril	—	ACE
10	Prinivil	lisinopril	ACE
11	Zestril	lisinopril	ACE

[a]Full classification list available in Multimedia Appendix 4.

[b]Generic drug name included for brand name drugs.

[c]Not applicable for generic drugs.

[d]ACE: angiotensin-converting enzyme inhibitor.

Drug Classification by RxNorm Concept Unique Identifier Utilizing the Drug Ontology

The initial DrOn RxCUI Classification contained 2543 antihypertensive RxCUIs that were mapped to 46 antihypertensive medication classes or combination medication classes. Each RxCUI entry contains the RxCUI, the drug product, and the drug class. An excerpt of the DrOn RxCUI Classification is shown in Table 3, and the full map is available in Multimedia Appendix 5.

Table 3. Excerpt from the Drug Ontology RxNorm Concept Unique Identifier classification.

RxCUI[a,b]	Drug_Product	Rx_Norm_Drug_Class
858926	Enalapril Maleate 20 MG Chewable Tablet	ACE[c]
378269	Enalapril Chewable Tablet	ACE
858810	Enalapril Maleate 20 MG Oral Tablet	ACE
858804	Enalapril Maleate 2.5 MG Oral Tablet	ACE
858821	Enalapril Maleate 1.25 MG/ML Injectable Solution	ACE
858817	Enalapril Maleate 10 MG Oral Tablet	ACE
858813	Enalapril Maleate 5 MG Oral Tablet	ACE
372007	Enalapril Oral Tablet	ACE
378288	Enalapril Injectable Solution	ACE
246264	Enalaprilat 1 MG/ML Injectable Solution	ACE
374378	Enalaprilat Injectable Solution	ACE
252820	Enalaprilat 0.625 MG/ML Injectable Solution	ACE
204404	Enalaprilat 1.25 MG/ML Injectable Solution	ACE

[a]Full map available in Multimedia Appendix 5.

[b]RxCUI: RxNorm Concept Unique Identifier.

[c]ACE: angiotensin-converting enzyme inhibitor.

Comparison Between Drug Classifications

The percent of drugs covered by each antihypertensive map is shown in Table 4. Using the Medication Name Classification, 207 ingredient terms were mapped to antihypertensive drug classes out of a total of 8025 unique raw medication names (2.58%; Table 4). When the DrOn RxCUI Classification was applied to the unique RxCUIs (n=13,627), 729 were successfully mapped to antihypertensive drug classes (5.35%; Table 4). Through the RxClass API, using ATC as the relationship source, 822 out of 13,602 RxCUIs were mapped to antihypertensive

drug classes (6.04%; Table 4). Finally, when the RxClass API was used with MED-RT as the relationship source, 792 RxCUIs had the relationship of "may_treat" HTN and a "has_MoA" that corresponded to an antihypertensive drug class (792/13,602, 5.82%; Table 4).

When the DrOn RxCUI Classification was compared with the other classifications, they broadly overlapped; however, there were some key differences (Figure 1). When the DrOn RxCUI Classification was compared with the other RxCUI-based maps constructed through the RxClass API, 691 terms overlapped with the RxClass-ATC map and 683 terms overlapped with the RxClass–MED-RT map. There were 38 terms and 46 terms that were unique to the DrOn RxCUI map when compared with the RxClass-ATC map and RxClass–MED-RT map, respectively. Additionally, there were 131 terms that were unique to the RxClass-ATC map, and 109 terms that were unique to the RxClass–MED-RT map. Of these 131 terms and 109 terms, respectively, there were 135 unique terms, with 105 overlapping between the RxClass-ATC map and the RxClass–MED-RT map.

Of the 135 unique RxCUIs from the RxClass ATC and MED-RT maps, 29 had a term type of *IN*, *MIN*, or *PIN*, representing antihypertensive medications. These term types were not included in the DrOn RxCUI map owing to their low level of specificity. Next, there were 24 RxCUIs that were either ophthalmic or topical drug products. Examples of these are shown in Table 5.

Table 4. Drug coverage by each classification method.

Classification method	Input term	Input, N	Antihypertensive	
			Mapped, n	Coverage, %
Medication Name	Raw Name	8025	207	2.58
DrOn[a] RxCUI[b]	RxCUI	13,627	729	5.35
RxClass-ATC[c]	RxCUI	13,602	822	6.04
RxClass–MED-RT[d,e]	RXCUI	13,602	792	5.82

[a]DrOn: Drug Ontology.

[b]RxCUI: RxNorm Concept Unique Identifier.

[c]ATC: Anatomical Therapeutic Chemical.

[d]MED-RT Antihypertensive coverage was mapped using 2 steps: (1) "may_treat" hypertension and (2) mechanism of action of an antihypertensive drug class.

[e]MED-RT: Medication Reference Terminology.

Figure 1. Comparisons of the Drug Ontology RxNorm Concept Unique Identifier Classification to the other classification methods. Results are shown for the number overlapping between the methods (Both) and the numbers unique to each method. ATC: Anatomical Therapeutic Chemical; MED-RT: Medication Reference Terminology.

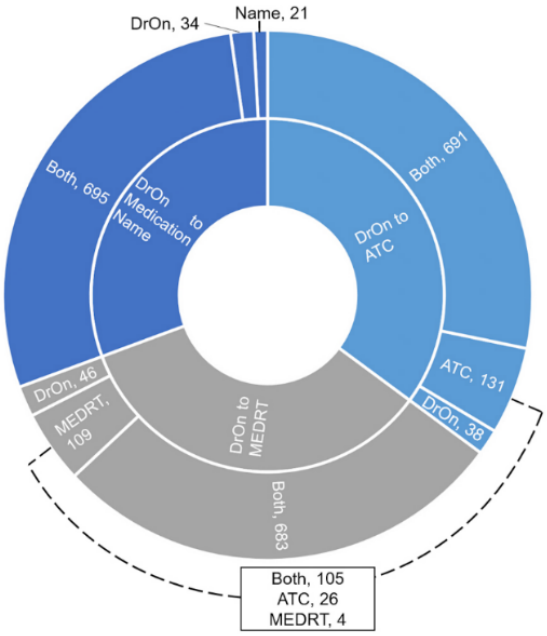


Table 5. Examples of RxNorm Concept Unique Identifiers inappropriately mapped as antihypertensives utilizing the RxMix tools.

RxCUI[a]	DrugName	hasMOA	ConceptName	DoseForm
207371	Minoxidil	Potassium Channel Interactions	Minoxidil 20 MG/ML Topical Solution (Rogaine)	Topical solution
208560	Timolol	Adrenergic beta1-Antagonists	Timolol 2.5 MG/ML Ophthalmic Solution (Betimol)	Ophthalmic solution
213729	Betaxolol	Adrenergic beta1-Antagonists	Betaxolol 2.5 MG/ML Ophthalmic Suspension (Betoptic S)	Ophthalmic suspension

[a]RxCUI: RxNorm Concept Unique Identifier.

Of the remaining 82 RxCUIs, 11 were RxCUIs corresponding to alpha-blockers primarily indicated for benign prostatic hyperplasia, 6 were RxCUIs for sacubitril/valsartan (Entresto) indicated for chronic heart failure with reduced ejection fraction, 1 was an RxCUI corresponding to an injectable antihypertensive product that is rarely used to treat HTN outside of a hypertensive crisis, and 1 was the *IN* term-type RxCUI for Potassium. A total of 63 were true antihypertensive drugs that were inadvertently omitted from the DrOn RxCUI map.

When the DrOn RxCUI Classification was compared with the Medication Name Classification using the DrOn RxCUI Classification as the reference, 695 terms overlapped, and 34 terms were unique to the DrOn RxCUI Classification (Figure 1). When the DrOn RxCUI Classification was compared with the Medication Name Classification using the Medication Name Classification as the reference, 195 terms overlapped, and 12 terms were unique to the Medication Name Classification. In total, 11 of the 12 terms that were unique to the Medication Name Classification were also included in the 135 terms that were unique to the RxClass ATC and MED-RT maps. The other term represented the brand name (Loniten) for the vasodilator minoxidil and was not present in any of the other classifications.

From the comparisons, the DrOn RxCUI Classification had 71 unique RxCUIs/terms that were not present in the RxClass MED-RT, RxClass ATC, and/or Medication Name classifications. Many were combination products (n=29), and the remaining 42 RxCUIs were spread across the other antihypertensive drug classes: ACE inhibitor (n=1), alpha-blocker (n=1), ARB (n=1), beta-blocker (n=17), CCB (n=11), loop diuretic (n=1), thiazide diuretic (n=3), and vasodilator (n=7).

Following the comparisons between the different drug classifications, 96 RxCUIs were added to the DrOn RxCUI Classification, and 15 antihypertensive medication names were added to the Medication Name Classification. The DrOn RxCUI Classification version 1.0 contains 2639 RxCUIs and is available in Multimedia Appendix 6, and the Medication Name Classification version 1.0 contains 301 antihypertensive medication names and is available in Multimedia Appendix 7 [10].

Application of the Drug Classifications to OneFlorida

The results from applying the DrOn RxCUI Classification v1.0 and the Medication Name Classification v1.0 to the prescribing information from the OneFlorida HTN population are shown in Table 6.

Overall, the methods performed very similarly, with approximately 15% (2,080,685 versus 2,089,557/13,879,046) of the total prescriptions mapping to antihypertensive drugs. The DrOn RxCUI Classification v1.0 was able to map 8872 more prescription records to an antihypertensive drug class compared with the Medication Name Classification v1.0 (Table 6). When the different methods were compared by antihypertensive class, all classes were within 1% of each other (eg, 443,835/2,089,557, 21.24% vs 443,540/2,080,685, 21.32% of prescriptions mapped to the ACE inhibitor class using the DrOn RxCUI Classification method and Medication Name Classification method, respectively). Table 7 shows the specific antihypertensive class or classes each prescription was mapped to using the DrOn RxCUI Classification v1.0. The majority of the prescriptions mapped to a single antihypertensive class (eg, ACE inhibitors, beta-bockers; Table 7). However, 10.42% (217,682/2,089,557) mapped to combination antihypertensive products (Table 7).

Table 6. Summary of the methods to the OneFlorida Antihypertensive Prescribing Dataset.

Classification method	Antihypertensive (N=13,879,046)	
	Mapped, n	Coverage, %
Medication Name	2,080,685	14.99
DrOn[a] RxCUI[b]	2,089,557	15.06

[a]DrOn: Drug Ontology.

[b]RxCUI: RxNorm Concept Unique Identifier.

Table 7. Application of the Drug Ontology RxNorm Concept Unique Identifier Classification to the OneFlorida antihypertensive prescribing dataset (N=2,089,557).

Antihypertensive Drug Class or Classes[a]	Frequency, n (%)
ACE[b]	443,835 (21.24)
Beta-blocker	411,721 (19.70)
Calcium Channel Blocker	360,653 (17.26)
Diuretic (Thiazide/Thiazide Like)	217,474 (10.41)
ARB[c]	178,252 (8.53)
Loop Diuretic	115,931 (5.55)
Diuretic/ACE Combo	92,275 (4.42)
Diuretic/ARB Combo	63,010 (3.02)
Centrally Acting	56,501 (2.70)
Vasodilator	30,665 (1.47)
Aldosterone Antagonist	29,214 (1.40)
Alpha Blocker	26,995 (1.29)

[a]Results are shown for antihypertensive drug classes that represent ≥1% of all antihypertensive prescriptions among the OneFlorida HTN population.

[b]ACE: angiotensin-converting enzyme inhibitor

[c]ARB: angiotensin II receptor inhibitor.

Discussion

We created 2 different medication classification systems for antihypertensive drugs: one utilizing medication names and the other utilizing RxCUIs. After comparing these classification systems to each other, and to existing drug class terminologies available through RxNorm, we identified key areas that can lead to misclassification of antihypertensive medications and drug classes. Most misclassifications stemmed from failure to discriminate between dosage forms or issues related to primary indications of drugs (eg, selection of drugs that are primarily indicated for benign prostatic hyperplasia).

To illustrate, timolol is a beta-blocker that has both oral and ophthalmic dosage forms. The oral form is used to treat HTN, whereas the ophthalmic form is used to treat glaucoma [9,24]. An ideal antihypertensive drug classification system would include the oral form of timolol, but not the ophthalmic form. This is exactly what we see with the DrOn RxCUI Classification, as DrOn represents the therapeutic indication (HTN) for oral timolol separate from its MoA (beta-blocker) [8]. In contrast, all forms of timolol were included in our analysis that utilized the ATC and MED-RT terminologies. ATC only modeled the drug class, with all forms of timolol included as *beta-blocking agents*. MED-RT allowed for filtering through both a MoA relationship and a may-treat relationship. However, all forms of timolol, both oral and ophthalmic are mapped back to the ingredient term type for timolol, which has the may-treat relationship with HTN. Although this method allows for simplicity in mapping multiple drug products, it assigns an incorrect therapeutic indication (for our purposes) to the ophthalmic forms of timolol. We observed similar misclassification for other medications that have multiple clinical uses beyond HTN (eg, the vasodilator minoxidil and the alpha-1-blockers silodosin, alfuzosin, and tamsulosin). Minoxidil has oral dosage forms to treat HTN and topical foams to treat hair loss [25,26], whereas silodosin, alfuzosin, and tamsulosin are all alpha-1-blockers that are indicated for the treatment of benign prostatic hyperplasia as opposed to HTN [27,28].

Retired RxCUIs, or those that have been remapped to other classes, were classified by DrOn but not by MED-RT or ATC. This illustrates the need for maps and terminologies that include retired and obsolete RxCUIs as many of our longitudinal data sources include these. The data source used in this study contains data from January 2011 to July 2017, and contained 1170 RxCUIs that have been retired and 421 that have been remapped.

We conducted this work in the context of optimizing antihypertensive medication classification for use in HTN-related computable phenotypes. In other disease states where there are not as many options for pharmaceutical treatment, the classification of drug products may not be as complicated. However, in the case of HTN, and particularly the complex clinical phenotype of resistant hypertension (RHTN) [29-31], there are multiple classes of antihypertensive drugs with different mechanisms of action [30,32]. With RHTN, it is necessary to properly assign each antihypertensive medication to a medication class to determine if a patient meets the definition for RHTN, which is classically defined as requiring 4 or more antihypertensive drugs from different antihypertensive drug classes to achieve BP control [30,31]. This also illustrates the need to include a subject matter expert (eg, PharmD) during the creation of drug classification systems.

We also observed differences in the classification of combination drug products. There were some combination antihypertensive drug products that were not identified in the MED-RT terminology through the methods that we utilized in this study. Additionally, when utilizing a classification system

based on medication name, all possible permutations of the combination name must be considered and included in the map (eg, HCTZ-metoprolol, hydrochlorothiazide-metoprolol, metoprolol-HCTZ, and metoprolol-hydrochlorothiazide). Without each of these permutations, it is possible to miss certain combination products. Finally, as a phenotype such as RHTN is determined partially, or fully, based on the antihypertensive drug count, a correct drug count must be assigned to each combination product. We have added this field into our DrOn RxCUI Classification v1.0.

Our study is not without limitation; currently, we do not have a gold standard antihypertensive medication classification list. We selected different data sources and compared classification based on these sources. However, we only used terminologies available through the US NLM. We have not included terminologies maintained by other groups (eg, American Hospital Formulary Service Pharmacologic-Therapeutic Classification) [33]. In addition, to classify based off of RxCUI, an RxCUI must be present. This may require additional work in some health care systems or datasets. We did not explore the temporal relationship between HTN diagnosis and antihypertensive medication prescription within our data. This could have resulted in the inclusion of medication data that was prescribed for indications other than HTN (eg, heart failure, atrial fibrillation, etc). We can examine this relationship more in our future work. Additionally, we did not explore the effect of the different methods and resources for creating the drug classification on the number of patients with RHTN within OneFlorida. We chose our study methodology because we first needed to create and validate this antihypertensive drug classification before we could use it in creating and validating our RHTN computable phenotype. Once our RHTN computable phenotype is validated, we can subsequently explore the effect of the different drug classification methods on the RHTN phenotype in future work. Other future work will include incorporating RxNorm Term Type (SCDF, SCD, SBD, etc), indicating retired RxCUIs, expanding to other disease states (diabetes, chronic kidney disease, heart failure, etc), and exploring the application to clinical decision support within the EHR [34,35].

In conclusion, we created and compared 4 different drug classification methods, focusing on the classification of antihypertensive drug products. We observed key differences in how each classification system handled drug products with multiple dosage forms, drug products with indications for benign prostatic hyperplasia, drug products that contain multiple antihypertensive ingredients (combination drug products), and RxCUIs that have been retired or remapped. We constructed 2 antihypertensive drug classification systems, 1 based off of RxCUIs and 1 based off of medication names. These are available for public use [10], and we will continue to update them through our own research.

Acknowledgments

Support for this project came from National Institutes of Health (NIH) grants KL2 TR001429 (CM), K01 HL141690 (CM), and K01 HL138172 (SS). Additionally, the research reported in this publication was supported in part by the OneFlorida Clinical Data Network, funded by the Patient-Centered Outcomes Research Institute (PCORI) #CDRN-1501-26692, in part by the OneFlorida Cancer Control Alliance, funded by the Florida Department of Health's James and Esther King Biomedical Research Program #4KB16, and in part by the University of Florida Clinical and Translational Science Institute, which is supported in part by the NIH National Center for Advancing Translational Sciences under award number UL1TR001427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the PCORI, its Board of Governors or Methodology, the OneFlorida Clinical Research Consortium, the University of Florida's Clinical and Translational Science Institute, the Florida Department of Health, or the NIH.

Conflicts of Interest

WRH is one of the creators of DrOn.

Multimedia Appendix 1
Features of each Classification Method.
[XLSX File (Microsoft Excel File), 9 KB - medinform_v8i2e14777_app1.xlsx]

Multimedia Appendix 2
List of Anatomical Therapeutic Chemical relationships included as antihypertensive drugs.
[XLSX File (Microsoft Excel File), 9 KB - medinform_v8i2e14777_app2.xlsx]

Multimedia Appendix 3
List of Medication Reference Terminology has_MoA relationships included as antihypertensive drugs.
[XLSX File (Microsoft Excel File), 9 KB - medinform_v8i2e14777_app3.xlsx]

Multimedia Appendix 4
Initial Medication Name Classification.

[XLSX File (Microsoft Excel File), 16 KB - medinform_v8i2e14777_app4.xlsx]

Multimedia Appendix 5
Initial Drug Ontology RxNorm Concept Unique Identifier Classification.
[XLSX File (Microsoft Excel File), 80 KB - medinform_v8i2e14777_app5.xlsx]

Multimedia Appendix 6
Drug Ontology RxNorm Concept Unique Identifier Classification version 1.0.
[XLSX File (Microsoft Excel File), 103 KB - medinform_v8i2e14777_app6.xlsx]

Multimedia Appendix 7
Medication Name Classification version 1.0.
[XLSX File (Microsoft Excel File), 17 KB - medinform_v8i2e14777_app7.xlsx]

References

1. Klann JG, Abend A, Raghavan VA, Mandl KD, Murphy SN. Data interchange using i2b2. J Am Med Inform Assoc 2016 Sep;23(5):909-915 [FREE Full text] [doi: 10.1093/jamia/ocv188] [Medline: 26911824]
2. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc 2014;21(4):578-582 [FREE Full text] [doi: 10.1136/amiajnl-2014-002747] [Medline: 24821743]
3. PCORnet. PCORnet. PCORnet Common Data Model (CDM) URL: https://pcornet.org/wp-content/uploads/2019/09/PCORnet-Common-Data-Model-v51-2019_09_12.pdf [accessed 2020-01-03]
4. Klann JG, Joss MA, Embree K, Murphy SN. Data model harmonization for the All Of Us Research Program: Transforming i2b2 data into the OMOP common data model. PLoS One 2019;14(2):e0212463 [FREE Full text] [doi: 10.1371/journal.pone.0212463] [Medline: 30779778]
5. Wei W, Denny JC. Extracting research-quality phenotypes from electronic health records to support precision medicine. Genome Med 2015;7(1):41 [FREE Full text] [doi: 10.1186/s13073-015-0166-y] [Medline: 25937834]
6. Newton KM, Peissig PL, Kho AN, Bielinski SJ, Berg RL, Choudhary V, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. J Am Med Inform Assoc 2013 Jun;20(e1):e147-e154 [FREE Full text] [doi: 10.1136/amiajnl-2012-000896] [Medline: 23531748]
7. Mo H, Thompson WK, Rasmussen LV, Pacheco JA, Jiang G, Kiefer R, et al. Desiderata for computable representations of electronic health records-driven phenotype algorithms. J Am Med Inform Assoc 2015 Nov;22(6):1220-1230 [FREE Full text] [doi: 10.1093/jamia/ocv112] [Medline: 26342218]
8. Hogan WR, Hanna J, Hicks A, Amirova S, Bramblett B, Diller M, et al. Therapeutic indications and other use-case-driven updates in the drug ontology: anti-malarials, anti-hypertensives, opioid analgesics, and a large term request. J Biomed Semantics 2017 Mar 3;8(1):10 [FREE Full text] [doi: 10.1186/s13326-017-0121-5] [Medline: 28253937]
9. Frampton JE, Perry CM. Topical dorzolamide 2%/timolol 0.5% ophthalmic solution: a review of its use in the treatment of glaucoma and ocular hypertension. Drugs Aging 2006;23(12):977-995. [doi: 10.2165/00002512-200623120-00005] [Medline: 17154662]
10. McDonough CW, Smith SM, Cooper-DeHoff RM, Hogan WR. GitHub. 2019. Antihypertensive Medication Classification (AntiHTNMedClassification) URL: https://github.com/caitrinmcd/AntiHTNMedClassification [accessed 2020-01-03]
11. Nelson SJ, Zeng K, Kilbourne J, Powell T, Moore R. Normalized names for clinical drugs: RxNorm at 6 years. J Am Med Inform Assoc 2011;18(4):441-448 [FREE Full text] [doi: 10.1136/amiajnl-2011-000116] [Medline: 21515544]
12. Bodenreider O, Cornet R, Vreeman DJ. Recent Developments in Clinical Terminologies - SNOMED CT, LOINC, and RxNorm. Yearb Med Inform 2018 Aug;27(1):129-139 [FREE Full text] [doi: 10.1055/s-0038-1667077] [Medline: 30157516]
13. Peters L, Bodenreider O. Using the RxNorm web services API for quality assurance purposes. AMIA Annu Symp Proc 2008 Nov;6:591-595 [FREE Full text] [Medline: 18999038]
14. RxNorm Navigator (RxNav). URL: https://rxnav.nlm.nih.gov/ [accessed 2019-12-18]
15. Bodenreider O, Rodriguez LM. Analyzing US prescription lists with RxNorm and the ATC/DDD Index. AMIA Annu Symp Proc 2014;2014:297-306 [FREE Full text] [Medline: 25954332]
16. Hogan WR, Hanna J, Joseph E, Brochhausen M. Towards a consistent and scientifically accurate drug ontology. CEUR Workshop Proc 2013;1060:68-73 [FREE Full text] [Medline: 27867326]
17. Hanna J, Joseph E, Brochhausen M, Hogan WR. Building a drug ontology based on RxNorm and other sources. J Biomed Semantics 2013 Dec 18;4(1):44 [FREE Full text] [doi: 10.1186/2041-1480-4-44] [Medline: 24345026]
18. Hanna J, Bian J, Hogan WR. An accurate and precise representation of drug ingredients. J Biomed Semantics 2016;7:7 [FREE Full text] [doi: 10.1186/s13326-016-0048-2] [Medline: 27096073]

19. Shenkman E, Hurt M, Hogan W, Carrasquillo O, Smith S, Brickman A, et al. OneFlorida Clinical Research Consortium: Linking a Clinical and Translational Science Institute With a Community-Based Distributive Medical Education Model. Acad Med 2018 Mar;93(3):451-455 [FREE Full text] [doi: 10.1097/ACM.0000000000002029] [Medline: 29045273]
20. Collins FS, Hudson KL, Briggs JP, Lauer MS. PCORnet: turning a dream into reality. J Am Med Inform Assoc 2014;21(4):576-577 [FREE Full text] [doi: 10.1136/amiajnl-2014-002864] [Medline: 24821744]
21. Sattari M, Cauley JA, Garvan C, Johnson KC, LaMonte MJ, Li W, et al. Osteoporosis in the Women's Health Initiative: Another Treatment Gap? Am J Med 2017 Aug;130(8):937-948. [doi: 10.1016/j.amjmed.2017.02.042] [Medline: 28366425]
22. Dumitrescu L, Ritchie MD, Denny JC, El Rouby NM, McDonough CW, Bradford Y, et al. Genome-wide study of resistant hypertension identified from electronic health records. PLoS One 2017;12(2):e0171745 [FREE Full text] [doi: 10.1371/journal.pone.0171745] [Medline: 28222112]
23. Medical Ontology Research. RxMix URL: https://mor.nlm.nih.gov/RxMix/ [accessed 2019-12-18]
24. Harris FJ, Tonkin M, Pratt C, DeMaria AN, Amsterdam EA, Mason DT. Short- and long-term therapy of mild essential hypertension with timolol. Clin Pharmacol Ther 1981 Dec;30(6):765-772. [doi: 10.1038/clpt.1981.236] [Medline: 7030579]
25. Pettinger WA. Minoxidil and the treatment of severe hypertension. N Engl J Med 1980 Oct 16;303(16):922-926. [doi: 10.1056/NEJM198010163031607] [Medline: 6997744]
26. Rumsfield JA, West DP, Fiedler-Weiss VC. Topical minoxidil therapy for hair regrowth. Clin Pharm 1987 May;6(5):386-392. [Medline: 3311578]
27. Jung JH, Kim J, MacDonald R, Reddy B, Kim MH, Dahm P. Silodosin for the treatment of lower urinary tract symptoms in men with benign prostatic hyperplasia. Cochrane Database Syst Rev 2017 Nov 22;11:CD012615 [FREE Full text] [doi: 10.1002/14651858.CD012615.pub2] [Medline: 29161773]
28. Maldonado-Ávila M, Manzanilla-García HA, Sierra-Ramírez JA, Carrillo-Ruiz JD, González-Valle JC, Rosas-Nava E, et al. A comparative study on the use of tamsulosin versus alfuzosin in spontaneous micturition recovery after transurethral catheter removal in patients with benign prostatic growth. Int Urol Nephrol 2014 Apr;46(4):687-690. [doi: 10.1007/s11255-013-0515-y] [Medline: 24061764]
29. Achelrod D, Wenzel U, Frey S. Systematic review and meta-analysis of the prevalence of resistant hypertension in treated hypertensive populations. Am J Hypertens 2015 Mar;28(3):355-361. [doi: 10.1093/ajh/hpu151] [Medline: 25156625]
30. Carey RM, Calhoun DA, Bakris GL, Brook RD, Daugherty SL, Dennison-Himmelfarb CR, American Heart Association Professional/Public Education and Publications Committee of the Council on Hypertension; Council on Cardiovascular and Stroke Nursing; Council on Clinical Cardiology; Council on Genomic and Precision Medicine; Council on Peripheral Vascular Disease; Council on Quality of Care and Outcomes Research;Stroke Council. Resistant Hypertension: Detection, Evaluation, and Management: A Scientific Statement From the American Heart Association. Hypertension 2018 Nov;72(5):e53-e90 [FREE Full text] [doi: 10.1161/HYP.0000000000000084] [Medline: 30354828]
31. Calhoun DA, Jones D, Textor S, Goff DC, Murphy TP, Toto RD, et al. Resistant hypertension: diagnosis, evaluation, and treatment. A scientific statement from the American Heart Association Professional Education Committee of the Council for High Blood Pressure Research. Hypertension 2008 Jun;51(6):1403-1419. [doi: 10.1161/HYPERTENSIONAHA.108.189141] [Medline: 18391085]
32. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Himmelfarb CD, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2018 Jun;71(6):1269-1324. [doi: 10.1161/HYP.0000000000000066] [Medline: 29133354]
33. American Society Of Health System Pharmacists (ASHP). AHFS Drug Information 2019. Bethesda, Maryland: Amer Soc Of Health System; 2019.
34. Young A, Tordoff J, Dovey S, Reith D, Lloyd H, Tilyard M, et al. Using an Electronic Decision Support Tool to Reduce Inappropriate Polypharmacy and Optimize Medicines: Rationale and Methods. JMIR Res Protoc 2016 Jun 10;5(2):e105 [FREE Full text] [doi: 10.2196/resprot.5543] [Medline: 27288200]
35. Dixon BE, Alzeer AH, Phillips EO, Marrero DG. Integration of provider, pharmacy, and patient-reported data to improve medication adherence for type 2 diabetes: a controlled before-after pilot study. JMIR Med Inform 2016 Feb 8;4(1):e4 [FREE Full text] [doi: 10.2196/medinform.4739] [Medline: 26858218]

Abbreviations

ACE: angiotensin-converting enzyme
API: application programming interface
ARB: angiotensin II receptor inhibitor
ATC: anatomical therapeutic chemical
BP: blood pressure
CCB: calcium channel blocker
CDM: common data model

DrOn: drug ontology
EHR: electronic health record
HTN: hypertension
ICD: International Classification of Diseases
MED-RT: Medication Reference Terminology
MoA: mechanism of action
NIH: National Institutes of Health
NLM: National Library of Medicine
PCORI: Patient-Centered Outcomes Research Institute
PCORnet: The National Patient Centered Clinical Research Network
RHTN: resistant hypertension
RxCUI: RxNorm Concept Unique Identifier

Edited by G Eysenbach; submitted 21.05.19; peer-reviewed by K Tingay, Y Chu; comments to author 29.09.19; revised version received 18.10.19; accepted 15.12.19; published 27.02.20.

Please cite as:
McDonough CW, Smith SM, Cooper-DeHoff RM, Hogan WR
Optimizing Antihypertensive Medication Classification in Electronic Health Record-Based Data: Classification System Development and Methodological Comparison
JMIR Med Inform 2020;8(2):e14777
URL: http://medinform.jmir.org/2020/2/e14777/
doi:10.2196/14777
PMID:32130152

Original Paper

Evaluation of Privacy Risks of Patients' Data in China: Case Study

Mengchun Gong[1*], MD; Shuang Wang[2*], PhD; Lezi Wang[1], MSc; Chao Liu[1], PhD; Jianyang Wang[3], MD; Qiang Guo[4], MSc; Hao Zheng[5], PhD; Kang Xie[6], PhD; Chenghong Wang[3], MSc; Zhouguang Hui[3,7], MD

[1]Digital China Health Technologies Corporation Limited, Beijing, China

[2]Shanghai Putuo People's Hospital, Tongji University, Shanghai, China

[3]Department of Radiation Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[4]Big Data Center, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[5]Department of Bioinformatics, Hangzhou Nuowei Information Technology, Hangzhou, China

[6]The Third Research Institute of Ministry of Public Security, Key Lab of Information Network Security, Ministry of Public Security, Shanghai, China

[7]Department of VIP Medical Services, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Zhouguang Hui, MD
Department of VIP Medical Services
National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College
Panjiayuan Nanli #17, Chaoyang District
Beijing, 100021
China
Phone: 86 010 87787656
Email: huizg@cicams.ac.cn

Abstract

Background: Patient privacy is a ubiquitous problem around the world. Many existing studies have demonstrated the potential privacy risks associated with sharing of biomedical data. Owing to the increasing need for data sharing and analysis, health care data privacy is drawing more attention. However, to better protect biomedical data privacy, it is essential to assess the privacy risk in the first place.

Objective: In China, there is no clear regulation for health systems to deidentify data. It is also not known whether a mechanism such as the Health Insurance Portability and Accountability Act (HIPAA) safe harbor policy will achieve sufficient protection. This study aimed to conduct a pilot study using patient data from Chinese hospitals to understand and quantify the privacy risks of Chinese patients.

Methods: We used g-distinct analysis to evaluate the reidentification risks with regard to the HIPAA safe harbor approach when applied to Chinese patients' data. More specifically, we estimated the risks based on the HIPAA safe harbor and limited dataset policies by assuming an attacker has background knowledge of the patient from the public domain.

Results: The experiments were conducted on 0.83 million patients (with data field of *date of birth, gender, and surrogate ZIP codes* generated based on home address) across 33 provincial-level administrative divisions in China. Under the Limited Dataset policy, 19.58% (163,262/833,235) of the population could be uniquely identifiable under the g-distinct metric (ie, 1-distinct). In contrast, the Safe Harbor policy is able to significantly reduce privacy risk, where only 0.072% (601/833,235) of individuals are uniquely identifiable, and the majority of the population is 3000 indistinguishable (ie the population is expected to share common attributes with 3000 or less people).

Conclusions: Through the experiments based on real-world patient data, this work illustrates that the results of g-distinct analysis about Chinese patient privacy risk are similar to those from a previous US study, in which data from different organizations/regions might be vulnerable to different reidentification risks under different policies. This work provides reference to Chinese health care entities for estimating patients' privacy risk during data sharing, which laid the foundation of privacy risk study about Chinese patients' data in the future.

(JMIR Med Inform 2020;8(2):e13046) doi:10.2196/13046

KEYWORDS

patient privacy; privacy risk; Chinese patients' data; data sharing; re-identification

Introduction

Background

Medical data are naturally distributed across institutions as patients might visit different hospitals at different times or for different diseases. To better understand the risk factors and efficacy of treatment, it is necessary to share data and analyze them. However, patient data are highly sensitive as they contain medical and personal identity information [1-5]. This is a ubiquitous problem. China has the largest population in the world, and the issue of privacy is becoming a big concern for the health care system to share medical data. Inappropriate handling of these sensitive data can lead to privacy leakage, which in turn can result in social embarrassment and commercial fraudulence [6-10].

In the United States, the Health Insurance Portability and Accountability Act (HIPAA) [11] safeguards the health care data. Thus, protected health information can only be considered as deidentified if it is sanitized by one of the following approaches specified by the HIPAA privacy rule [12]: (1) expert determination or (2) safe harbor. The first approach is to recruit an expert with appropriate knowledge and experience to render information with minimal risk to be reidentified. The second approach is to use the safe harbor approach, which explicitly denotes 18 identifiers that need to be removed. The average fine levied for a HIPAA breach is between US $10,000 and US $50,000 per medical record. There are similar guidelines in other countries, for example, the European Union's General Data Protection Regulation [13] and Canada's Personal Information Protection and Electronic Documents Act [14], which regulate patient records and other sensitive information. In South Korea (Korea) and Japan, the general law regulating privacy and data protection is the Personal Information Protection Act, and there is a more complete list of international privacy-related laws by country and region.

In China, the *Network Security Law of the People's Republic of China* [15], which was formally put into effect on June 1, 2017, regulates that network providers must not disclose, falsify, or destroy any personal information they have collected. Any network provider must not disclose this personal information to any third party without obtaining consent from data owners, except for the data that cannot be used to reidentify a specific individual. However, there are no guidelines on how personal information can be processed to satisfy the above regulation. On December 29, 2017, the Chinese government formally released a new regulation called *Information Security Technology and Personal Information Security Specification* (referred to as *Specification*) [16]. In the *Specification*, the Chinese government has clearly defined privacy-related terms such as "personal information controller," "collection," "informed consent," "user portrait," "personal information security impact assessment," "deletion," and "deidentification." The *Specification* also defines security requirements for different phases (eg, collection, storage, processing, transfer, and disclosure) in handling personal data. However, the *Specification* also has several limitations. First, the *Specification* is a recommended national standard and not a legal regulation; thus, it might not be stringently enforced by different entities. Second, the *Specification* mainly focuses on general purpose information security, where no specific guidance is provided for tackling medical or health care data. For example, in the *Specification*, almost all medical-related data are defined as highly sensitive data. The *Specification*, on one hand, emphasizes the importance of obtaining explicit consent of individuals when collecting, using, or disclosing sensitive personal information, whereas, on the other hand, there are several situations have been added as exceptions. For instance, if the personal information controller is an academic research institution, then it is necessary for them to perform statistical or academic research in public interest. If they provide external academic research or description results with deidentified personal information, they will be exempted from obtaining explicit consent from each individual. In addition, if the use of personal data is directly related to public safety, public health, and major public interest, then there is no need to obtain individual consent on personal data usage. In the third case, if there are certain difficulties in obtaining personal consent and if the use of personal data is to safeguard the major legal rights such as the life and property of the subject or individuals, then such usage of personal information will be exempted from obtaining explicit consent. In the *Specification*, deidentification is defined as a process by which the personal information is technically marked out so that the remaining information cannot be used to reidentify the individual without using additional information. On August 15, 2017, *Information Security Technology and Personal Information Deidentification Specification* was published by the Chinese government for public comments, which also introduced many existing deidentification procedures. However, there is still no clear guidance in China about how to deidentify health care data to ensure sufficient protection of the privacy of individual patients. Owing to the difference in population density, it is also not clear if similar protection mechanisms such as the HIPAA safe harbor rule will provide comparable protection to the Chinese patients' data. There is also a difference between the external sources of background information that attackers can leverage. For example, there is no public voter's registry in China, but social networks make public a considerable amount of demographic information of their users such as gender, birthday, school, and job. It is necessary to measure the privacy risks of Chinese patients' data to better understand the associated privacy risk.

Besides direct identifiers (such as name, national ID, and address), the privacy risk of a medical record is related to the rareness of its key variable values. For example, if there is a unique combination of birthday, gender, and ZIP code, the corresponding record is more likely to be reidentified when compared with records that have duplicated characteristics in the database. It has been shown in the study by Sweeney [17]

that 87% of the US population can be uniquely identified by the triplet (birthday, gender, and ZIP code), which reveals a high privacy risk if data are shared without sanitization. It is important to measure the rareness of individual records in a database to understand the potential risk it carries.

Privacy risk measurements and anonymization methods such as *k*-anonymity [18], *l*-diversity [19], *t*-closeness [20], and differential privacy (DP) [21] have motivated many algorithmic and theoretical studies. *k*-anonymity reduces the granularity of data representation using data generalization and suppression technologies. The parameter *k* indicates the number of records within the equivalence class, in which an adversary cannot distinguish an individual. A larger *k* implies a smaller reidentification risk. El Emam et al [22] applied an optimized *k*-anonymity algorithm for health data deidentification. *l*-diversity improves *k*-anonymity by ensuring that the intragroup diversity for sensitive values is controlled by the parameter *l* [19]. *t*-closeness provides a stronger privacy notion than *l*-diversity, where *t*-closeness requires that the distribution of a private attribute in any equivalence class is computationally indistinguishable (ie, no larger than *t*) from the distribution of the attribute in the overall table. Both techniques have been adopted in many medical data deidentification applications [23]. Recently, DP became one of the de facto standards for achieving strong privacy guarantees, which assumes that an attacker with any background knowledge cannot tell if a particular individual's information has been included or not based on the differentially private outputs [24]. DP technology has also been applied to protect health care data dissemination and analysis [25-27]. In this work, we were interested in a measurement to evaluate the reidentification risks with respect to the HIPAA privacy rule when applied to Chinese patients' data. However, none of the aforementioned methods can be directly adopted for serving this goal. Therefore, this work resorts to the g-distinct method previously proposed in the study by Malin et al [28] for evaluating reidentification risks of HIPAA-deidentified data in the United States.

Objectives

The main objectives and contributions of this work are three-fold: (1) to provide one of the first large-scale studies on the privacy risks of Chinese patients' data, (2) to design specifically experimental studies based on the characteristics of Chinese patients' data for evaluating the patient privacy risks in China, and (3) to provide references for improving the current privacy protection and rulemaking for Chinese patients' data.

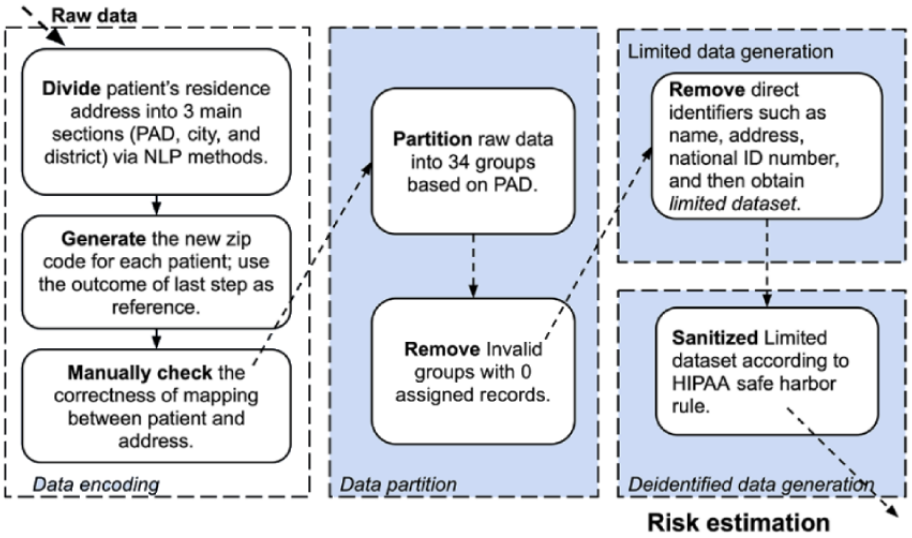
Methods

Data Preprocessing

The datasets used for conducting our experiments are based on cancer patients' records in the Malignant Cancer Big-data Processing Analysis and Application Research Project (MCBPAARP), which is supported by the National Cancer Institute's High-Tech Research and Development *863* program. The *863* program was led by the Ministry of Science and Technology of the People's Republic of China, where the goal of this program is to promote the development of advanced technologies across different fields. This study under MCBPAARP has been approved by the Institutional Review Board of National Cancer Center/National Cancer Hospital and Chinese Academy of Medical Sciences, also known as Ethics Committee of National Cancer Center, National Cancer Hospital, and Chinese Academy of Medical Sciences of Peking Union Medical College under the project ID 2017YFC1311000. In China, hospitals have to update inpatients' medical record home page to National Health Commission of the People's Republic of China under a unified standard. The Chinese patients' data attributes used in this study include fields P5 gender, P6 birthday, and P801 home address (used to generate masked ZIP codes).

Figure 1 illustrates the methods used for the raw data preprocessing in this study for privacy risk analysis, which includes four phases: (1) data encoding; (2) data partition; (3) limited dataset generation; and (4) *deidentified* dataset generation. The phases are described as follows:

Figure 1. The workflow of raw data preprocessing in this study. HIPAA: Health Insurance Portability and Accountability Act; NLP: neurolinguistic programming; PAD: provincial-level administrative division.



Data Encoding

In Chinese patients' data, the quality of ZIP codes from patients' raw data is extremely low, which may be either missing or too generalized (ie, only at city level). To overcome this problem, we introduced the following encoding scheme to convert the patient's address information into geocodes as surrogate ZIP codes in this study. We first divided the patient's residence address into three sections (ie, provincial-level administrative divisions [PADs], city, and district) by using natural language processing methods. Thereafter, we encoded PAD, city, and district with 2 digits, 3 digits, and 4 digits, respectively, which resulted in surrogate ZIP codes for a total of 9 digits. To ensure the data quality, we conducted two rounds of manual checking for the mapping correctness between the patient's residence addresses and their surrogate ZIP codes. We excluded patients with missing residence address information and the records with obvious logical error (ie, the patient's date of birth [DOB] is 1900-01-01). Finally, 0.83 million hospitalized patients' medical records were selected in this study.

Data Partition

We partitioned raw patient data into different groups based on their PADs. Through this phase, we ended up with 33 nonempty PADs except for Hong Kong PAD (see Multimedia Appendix 1 for more details of PADs).

Limited Dataset Generation

After the data encoding phase, we further removed additional explicit identifiers, such as name, address, and national ID number from the raw data, which left us with the *limited dataset* with only DOB, gender, and surrogate ZIP codes.

Deidentified Dataset Generation

On the basis of the HIPAA safe harbor rule, we further sanitized the limited dataset by generalizing DOB to year and all surrogate ZIP codes to the first 6 digits.

Risk Evaluation

To evaluate the privacy risk of the preprocessed Chinese patients' data, we adopted the g-distinct method introduced in the study by Malin et al [28] for studying a similar problem in the United States. The g-distinct method quantifies the uniqueness of individual records within a database, where an individual is said to be unique if such an individual has a combination of personal attributes that no other individuals in the same dataset has. Furthermore, we say an individual is g-distinct if the combination of their attributes is identical to at the most g-1 other individuals in the whole dataset space. For example, suppose an individual has the following combination of attributes: age at 35 years and gender as male. If there does not exist any other individual whose age and gender are also 35 years and male, respectively, then such an individual is considered as unique (ie, 1-distinct). In addition, if the total number of individuals with the same combination of attributes is equal to *k*, then we state this individual is *k-distinct*.

In other words, we can also describe the g-distinct as the sum over the number of patients in all bins with less than or equal to *g* individuals, which can be written as shown in equation (1):

$$h(g)=\sum_{i=1}^{g}i|bin(i)| \quad (1)$$

Here *g* denotes the parameter and |bin(i)| refers to the number of bins with exact *i* patients having identical attributes. This measurement serves as a proxy to the risk of stratified population with different combinations of characteristics. In this study, we applied the above *g*-distinct metric to the Chinese patients' data to evaluate the privacy risk.

Results

Experimental Setup

The g-distinct analysis is a population inspection method that allows us to investigate a particular cross-section for specific population collection. Such particular cross-section represents the set of individuals whose private records are most vulnerable to reidentification attacks. In our experiments, we conducted g-distinct analysis over the limited dataset (ie, DOB, gender, and ZIP code) and the safe harbor dataset (ie, birth year, gender, and masked ZIP code) to examine how the safe harbor data can improve the privacy of individual patients over limited data.

Experimental Results

The results of g-distinct analysis based on nationwide datasets for both the safe harbor dataset and limited dataset are illustrated in Figure 2. In Figure 2, the left and right subgraphs represent the g-distinct analysis results for limited and safe harbor datasets, respectively. According to the nationwide g-distinct analysis results, we have two major observations. On the one hand, without sufficient deidentification process (ie, the limited data on the left), the whole dataset is highly risky. For instance, 19.58% (163,262/833,235) of the population is 1-distinct in the limited dataset (ie, uniquely identifiable under the g-distinct metric). In addition, more than 90.6% of the population is 10-distinct, which implies that the majority of the population in the limited dataset is expected to share common attributes with 10 or less people. Such sheer number of distinct individuals results in a huge difficulty for privacy protection. Thus, in such cases, the limited data are extremely vulnerable to reidentification attacks. On the other hand, as shown in Figure 2, the safe harbor dataset is able to significantly preserve the patient's privacy, in which only 0.072% (601/833,235) of individuals are uniquely identifiable (ie, 1-distinct), and the majority of the population (around 95%) is 3000 indistinguishable.

We also studied the relationship between distinct individuals and the underlying populations, which simulates the impact of different ZIP code–masking strategies on the privacy protection. The results of this experiment have been illustrated in Figure 3. There are a total of 34 PADs in China. As there were no patient records with residence address within Hong Kong in the collected datasets, we estimated the percentage of 1-distinct population over the other 33 PADs (see Multimedia Appendix 2 for more details). Figure 3 shows the percentage of 1-distinct population associated with each PAD in an ascending order of the sample population in the given PAD. The 2 subgraphs are the results over limited dataset and the safe harbor dataset, respectively. Owing to the accommodation of different scales of 1-distinct percentage along with the increase in population,

the 2 plots are depicted in log-log scale. As shown in Figure 3, both results show a similar tendency, where the percentage of 1-distinct population decreases as the sampled population increases in different PADs. This is because a PAD with more sampled population tends to result in a higher probability to have more than 1 patient who shares the same attributes. Another observation from the result is that when the sampled population increases, the percentage of 1-distinct population of the safe harbor dataset decreases dramatically. When the population has increased to 10,000, the 1-distinct percentage has already dropped to 0.05%. In contrast, the decreasing tendency for the limited dataset seems more moderate (ie, potentially higher privacy risks). We can see that there is still 5% population more with 1-distinct for a PAD of 200,000 population in the limited dataset.

Figure 2. The g-distinct versus percentage of population under limited dataset and safe harbor dataset, respectively.

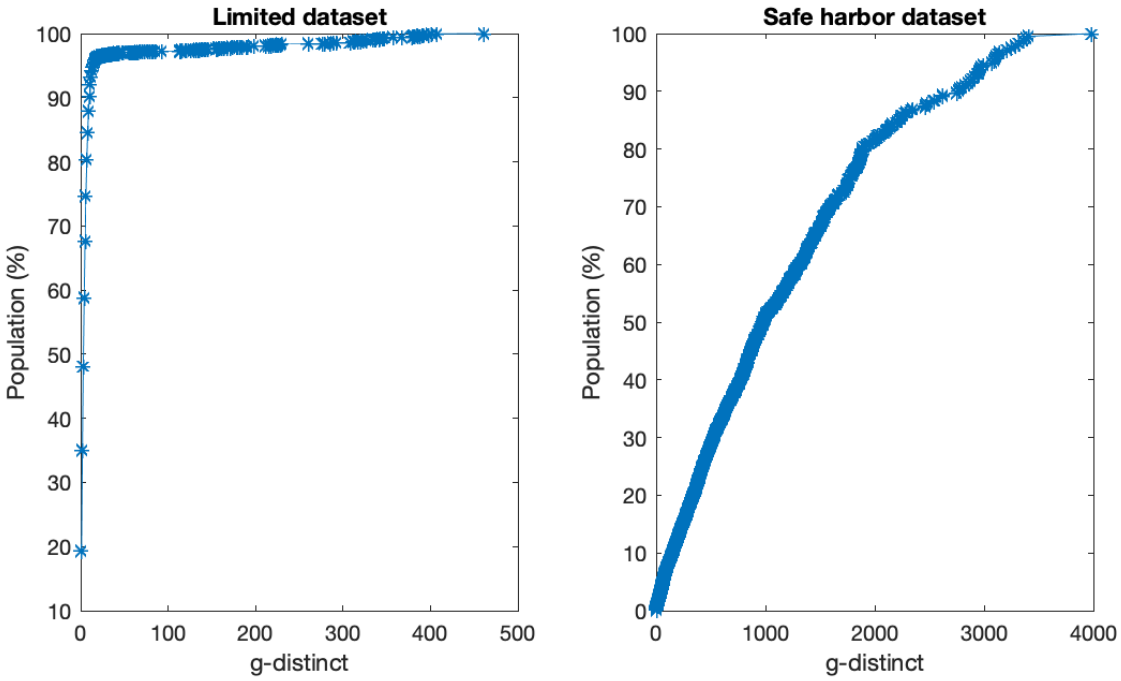
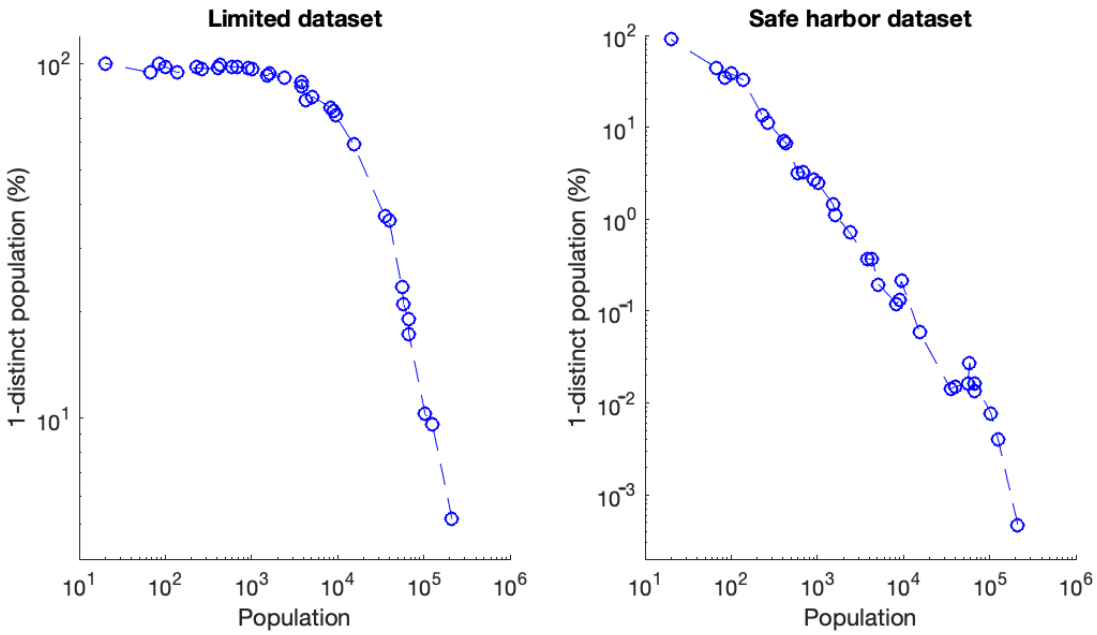


Figure 3. Percentage of 1-distinct versus total population under limited dataset and safe harbor dataset, respectively.



Discussion

Principal Findings

This work is one of the first large-scale studies on evaluating privacy risks of Chinese patients' data, which analyzed the reidentification risks based on the HIPAA safe harbor and limited dataset policies. The originality of this work can be summarized in three ways:

1. Originality in exploring new observations: Although many Chinese Acts [15] and national specification/regulations [16] have cited the HIPAA safe harbor rule [11] as a reference standard for guiding patient data deidentification in China, there is still a lack of quantitative observation on the reidentification risk of Chinese patient data when applying the HIPAA safe harbor standard. This work provides one of the first quantitative studies on large-scale nationwide Chinese patients' data toward this goal.
2. Originality in designing new experiments: Some Chinese patients' data attributes are unique and different from those in the United States. Therefore, these data cannot be applied directly to the risk assessment method used in previous US studies. For example, Chinese patients' data typically have extremely low quality of ZIP codes, which may be either missing or too generalized (ie, only at city level). Thus, we designed new data encoding, data partition, and data masking schemes based on Chinese patients' data characteristics to meet this goal.
3. Originality in contributing new knowledge: We made an assumption that the risk evaluation scheme defined by the HIPAA is satisfactory with respect to Chinese patient data as well. According to this assumption, we designed and implemented our experimental studies based on Chinese patients' data. As patient privacy protection is a very important topic, many other research studies have been conducted in Europe [29-31], Japan [32,33], and Australia [34]. However, most of these studies are more qualitative in orientation and usually not suitable for Chinese patients' data, which mainly focus on the interpretation and comparison of laws and regulations. In contrast, the focus of this work was to quantify the Chinese patient privacy risks with large-scale and real-world patient data collected from China. According to our experimental studies, our assumption is supported by the results of this work, which illustrates findings similar to those of a previous US study by Malin et al [28] that evaluated reidentification of US patient data associated with the HIPAA policies. Such studies are amenable to various kinds of meta-evaluations, enabling administrative roles such as government's policy makers and datacenter administrators to be able to evaluate policies and to determine the potential impact on reidentification risk. The experimental results demonstrate the power of the g-distinct analysis applied on Chinese patients' data. In general, according to the experimental results, the safe harbor dataset provides much stronger privacy protection in terms of 1-distinct than that provided by the limited dataset in Chinese patients' data.

Limitations

In general, this work provides justification for reidentification risk estimates on Chinese patient records before sharing data. However, the proposed studies still have a few limitations. First, the privacy risk that we estimated for the case study is based on the cancer patients' data without including patients with other diseases. Second, although the datasets are from 33 of 34 PADs, the study is still limited by the data scale, which only covers less than 0.06% of the Chinese population (ie, 0.83 million patients' records vs 1.5 billion total population). Third, the demographic information used in this study is also limited. For example, it is unfeasible to measure the identifiability based on nationality. Therefore, raw data are collected with selection bias because of aforementioned limitations. All these limitations justify further investigation along this line.

Conclusions

The study of Chinese patients' privacy risk in this work fills the gap of the privacy research between the United States and China. Moreover, as the Chinese government has not yet issued specific regulations or policies directly against privacy protection of citizens' health data, our experimental studies have the potential for Chinese officials to improve current health data–sharing regulations. The policy might vary largely among provinces, as according to the statistics, the g-distinct measurements vary widely across the provinces as well. Privacy officials might issue flexible policies for different regions.

Acknowledgments

This study is supported by a national key research and development program (2017YFC1311000, 2017YFC1311001). This work is partially supported by Key Lab of Information Network Security of Ministry of Public Security (The Third Research Institute of Ministry of Public Security) under funding C19609. The authors would like to thank Dr Hua Xu and Dr Xiaoqian Jiang for the helpful discussions and suggestions.

Authors' Contributions

MG and SW share first authorship and contributed the majority of the writing and conducted major parts of the methods and experiment design, where a part of this work was done by SW in the Institutes for Systems Genetics West China Hospital. LW and CL conducted some experiments and contributed to data preprocessing and paper writing. HZ and KX contributed to paper writing and provided comments on methods. JW, QG, and ZH provided the motivation for this work, data collection, paper writing, detailed edits, and critical suggestions. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Abbreviations of provinces in China.
[DOCX File , 15 KB - medinform_v8i2e13046_app1.docx]

Multimedia Appendix 2
The g-distinct analysis results.
[DOCX File , 18 KB - medinform_v8i2e13046_app2.docx]

References

1. Jiang Y, Hamer J, Wang C, Jiang X, Kim M, Song Y, et al. SecureLR: Secure Logistic Regression Model via a Hybrid Cryptographic Protocol. IEEE/ACM Trans. Comput. Biol. and Bioinf 2019 Jan 1;16(1):113-123. [doi: 10.1109/tcbb.2018.2833463]
2. Chen F, Wang C, Dai W, Jiang X, Mohammed N, Al Aziz MM, et al. PRESAGE: PRivacy-preserving gEnetic testing via SoftwAre Guard Extension. BMC Med Genomics 2017 Jul 26;10(S2). [doi: 10.1186/s12920-017-0281-2]
3. Wang X, Tang H, Wang S, Jiang X, Wang W, Bu D, et al. iDASH secure genome analysis competition 2017. BMC Med Genomics 2018 Oct 11;11(S4). [doi: 10.1186/s12920-018-0396-0]
4. Chenghong W, Jiang Y, Mohammed N, Chen F, Jiang X, Al Aziz MM, et al. SCOTCH: Secure Counting Of encrypTed genomiC data using a Hybrid approach. AMIA Annu Symp Proc 2017;2017:1744-1753 [FREE Full text] [Medline: 29854245]
5. Chen F, Wang S, Jiang X, Ding S, Lu Y, Kim J, et al. PRINCESS: Privacy-protecting Rare disease International Network Collaboration via Encryption through Software guard extensionS. Bioinformatics 2017 Mar 15;33(6):871-878 [FREE Full text] [doi: 10.1093/bioinformatics/btw758] [Medline: 28065902]
6. Krishnamurthy B, Wills CE. Privacy Leakage in Mobile Online Social Networks. 2010 Jun 10 Presented at: The 3rd Wonference on Online Social Networks; 2010; Berkeley, California URL: https://www.usenix.org/legacy/events/wosn10/tech/full_papers/Krishnamurthy.pdf
7. Ignatenko T, Willems FMJ. Privacy leakage in biometric secrecy systems. 2008 Jun 10 Presented at: 46th Annual Allerton Conference on Communication, Control, and Computing; 2008; Allerton. [doi: 10.1109/ALLERTON.2008.4797752]
8. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science 2013 Jan 18;339(6117):321-324 [FREE Full text] [doi: 10.1126/science.1229566] [Medline: 23329047]
9. Hansson MG, Lochmüller H, Riess O, Schaefer F, Orth M, Rubinstein Y, et al. The risk of re-identification versus the need to identify individuals in rare disease research. Eur J Hum Genet 2016 Nov 25;24(11):1553-1558 [FREE Full text] [doi: 10.1038/ejhg.2016.52] [Medline: 27222291]
10. Vaidya J, Shafiq B, Jiang X, Ohno-Machado L. Identifying inference attacks against healthcare data repositories. AMIA Jt Summits Transl Sci Proc 2013;2013:262-266 [FREE Full text] [Medline: 24303279]
11. Public Law. 1996. Health insurance portability and accountability act of 1996 URL: https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf [accessed 2020-01-17]
12. Summary of the HIPAA privacy rule. 2003. Others URL: https://www.hhs.gov/sites/default/files/privacysummary.pdf [accessed 2020-01-17]
13. Voigt P, von dem Bussche A. The EU General Data Protection Regulation (GDPR): A Practical Guide. Cham, Switzerland: Springer; Jun 10, 2017.
14. Charnetski W, Flaherty P, Robinson J. The Personal Information Protection and Electronic Documents Act: A Comprehensive Guide. In: Canada Law Book. Toronto: Canada Law Book; Jun 10, 2001:2001.
15. Peng PL. Ordinance of the People's Republic of China on the Protection of Computer Information System Security. Chinese Law & Government 2014 Dec 07;43(5):12-16. [doi: 10.2753/clg0009-4609430501]
16. China Law Blog. 2018. China’s Personal Information Security Specification: Get Ready for May 1 URL: https://www.chinalawblog.com/2018/02/chinas-personal-information-security-specification-get-ready-for-may-1.html [accessed 2018-06-26]
17. Sweeney L. CiNii. 2000. Uniqueness of simple demographics in the US Population, in LIDAP-WP4 URL: https://ci.nii.ac.jp/naid/10020493621/ [accessed 2020-01-17]
18. Sweeney L. Achieving k-Anonymity Privacy Protection Using Generalization and Suppression. Int J Unc Fuzz Knowl Based Syst 2012 May 02;10(05):571-588 [FREE Full text] [doi: 10.1142/S021848850200165X]
19. Machanavajjhala A, Gehrke J, Kifer D. L-diversity: Privacy beyond k-anonymity. 2006 Presented at: 22nd International Conference on Data Engineering (ICDE'06); April 3-7, 2006; Atlanta, GA. [doi: 10.1109/icde.2006.1]

20. Li N, Li T, Venkatasubramanian S. t-closeness: Privacy beyond k-anonymity and l-diversity. 2007 Presented at: 2007 IEEE 23rd International Conference on Data Engineering; June 4, 2007; Istanbul, Turkey URL: http://ieeexplore.ieee.org/abstract/document/4221659/ [doi: 10.1109/icde.2007.367856]
21. Dwork C. Differential Privacy: A Survey of Results. In: Theory and Applications of Models of Computation. Berlin, Heidelberg, Germany: Springer; Jun 10, 2008:19.
22. El Emam K, Dankar F, Issa R. A globally optimal k-anonymity method for the de-identification of health data. J Am Med Inform Assoc 2009;16:82. [doi: 10.1197/jamia.m3144]
23. Gkoulalas-Divanis A, Loukides G, Sun J. Publishing data from electronic health records while preserving privacy: a survey of algorithms. J Biomed Inform 2014 Aug;50:4-19 [FREE Full text] [doi: 10.1016/j.jbi.2014.06.002] [Medline: 24936746]
24. Jiang X, Sarwate AD, Ohno-Machado L. Privacy Technology to Support Data Sharing for Comparative Effectiveness Research. Medical Care 2013;51:S58-S65. [doi: 10.1097/mlr.0b013e31829b1d10]
25. Jiang Y, Wang C, Wu Z, Du X, Wang S. Privacy-preserving biomedical data dissemination via a hybrid approach. AMIA Annu Symp Proc 2018;2018:1176-1185 [FREE Full text] [Medline: 30815160]
26. 26 DF, El EK. The application of differential privacy to health data. 2012 Presented at: Proceedings of the Joint EDBT/ICDT Workshops; March 2012; Berlin, Germany URL: https://dl.acm.org/citation.cfm?id=2320816 [doi: 10.1145/2320765.2320816]
27. Dankar F, El EK. Practicing differential privacy in health care: A review. Trans Data Priv 2013;6:67.
28. Benitez K, Malin B. Evaluating re-identification risks with respect to the HIPAA privacy rule. J Am Med Inform Assoc 2010;17(2):169-177 [FREE Full text] [doi: 10.1136/jamia.2009.000026] [Medline: 20190059]
29. Tucker K, Branson J, Dilleen M, Hollis S, Loughlin P, Nixon MJ, et al. Protecting patient privacy when sharing patient-level data from clinical trials. BMC Med Res Methodol 2016 Jul 08;16 Suppl 1(S1):77 [FREE Full text] [doi: 10.1186/s12874-016-0169-4] [Medline: 27410040]
30. Wierda E, Eindhoven D, Schalij M, Borleffs CJW, Amoroso G, van Veghel D, et al. Privacy of patient data in quality-of-care registries in cardiology and cardiothoracic surgery: the impact of the new general data protection regulation EU-law. Eur Heart J Qual Care Clin Outcomes 2018 Oct 01;4(4):239-245. [doi: 10.1093/ehjqcco/qcy034] [Medline: 30060178]
31. Deguara I. Thesynapse.net. 2018. Protecting patients' medical records under the GDPR URL: https://www.um.edu.mt/library/oar/bitstream/123456789/40287/1/The_Synapse%2c_17%282%29_-_A1.pdf [accessed 2020-01-17]
32. Kim J. J Marshall J Info Tech & Privacy L. 2015. Japanese and American Privacy Laws, Comparative Analysis URL: https://repository.jmls.edu/cgi/viewcontent.cgi?article=1782&context=jitpl [accessed 2020-01-17]
33. Yamamoto H. Use of personal information in medical research in Japan. The Lancet 2016 Oct;388(10055):1981-1982. [doi: 10.1016/s0140-6736(16)31867-0]
34. Williamson OD, Cameron PA, McNeil JJ. Medical registry governance and patient privacy. Medical Journal of Australia 2004 Aug 02;181(3):125-126. [doi: 10.5694/j.1326-5377.2004.tb06200.x]

Abbreviations

DOB: date of birth
DP: differential privacy
HIPAA: Health Insurance Portability and Accountability Act
MCBPAARP: Malignant Cancer Big-data Processing Analysis and Application Research Project
PAD: provincial-level administrative division

Edited by G Eysenbach; submitted 15.12.18; peer-reviewed by J Bian, YR Park, A Alaqra; comments to author 17.04.19; revised version received 07.06.19; accepted 26.09.19; published 05.02.20.

Please cite as:
Gong M, Wang S, Wang L, Liu C, Wang J, Guo Q, Zheng H, Xie K, Wang C, Hui Z
Evaluation of Privacy Risks of Patients' Data in China: Case Study
JMIR Med Inform 2020;8(2):e13046
URL: https://medinform.jmir.org/2020/2/e13046
doi:10.2196/13046
PMID:32022691

Original Paper

Identifying Acute Low Back Pain Episodes in Primary Care Practice From Clinical Notes: Observational Study

Riccardo Miotto[1,2,3], PhD; Bethany L Percha[2,3], PhD; Benjamin S Glicksberg[1,2,3], PhD; Hao-Chih Lee[2,3], PhD; Lisanne Cruz[4], MD, MSc, FAAPMR; Joel T Dudley[2,3], PhD; Ismail Nabeel[5], MD, MPH

[1]Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[2]Institute for Next Generation Healthcare, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[4]Department of Physical Medicine and Rehabilitation, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[5]Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, United States

Corresponding Author:
Ismail Nabeel, MD, MPH
Department of Environmental Medicine and Public Health
Icahn School of Medicine at Mount Sinai
17 East 102nd Street, Box 1043
New York, NY, 10029
United States
Phone: 1 (614) 423 9057
Email: ismail.nabeel@icahn.mssm.edu

Abstract

Background: Acute and chronic low back pain (LBP) are different conditions with different treatments. However, they are coded in electronic health records with the same International Classification of Diseases, 10th revision (ICD-10) code (M54.5) and can be differentiated only by retrospective chart reviews. This prevents an efficient definition of data-driven guidelines for billing and therapy recommendations, such as return-to-work options.

Objective: The objective of this study was to evaluate the feasibility of automatically distinguishing acute LBP episodes by analyzing free-text clinical notes.

Methods: We used a dataset of 17,409 clinical notes from different primary care practices; of these, 891 documents were manually annotated as *acute LBP* and 2973 were generally associated with LBP via the recorded ICD-10 code. We compared different supervised and unsupervised strategies for automated identification: keyword search, topic modeling, logistic regression with bag of n-grams and manual features, and deep learning (a convolutional neural network-based architecture [ConvNet]). We trained the supervised models using either manual annotations or ICD-10 codes as positive labels.

Results: ConvNet trained using manual annotations obtained the best results with an area under the receiver operating characteristic curve of 0.98 and an F score of 0.70. ConvNet's results were also robust to reduction of the number of manually annotated documents. In the absence of manual annotations, topic models performed better than methods trained using ICD-10 codes, which were unsatisfactory for identifying LBP acuity.

Conclusions: This study uses clinical notes to delineate a potential path toward systematic learning of therapeutic strategies, billing guidelines, and management options for acute LBP at the point of care.

(JMIR Med Inform 2020;8(2):e16878) doi:10.2196/16878

KEYWORDS

electronic health records; clinical notes; low back pain; natural language processing; machine learning

Introduction

Low back pain (LBP) is one of the most common causes of disability in US adults younger than 45 years [1], with 10 to 20% of American workers reporting persistent back pain [2]. LBP impacts one's ability to work and affects the quality of life. For example, in 2015, Luckhaupt et al showed that, from a pool of 19,441 people, 16.9% of workers with any LBP and 19.0% of those with frequent and severe LBP missed at least one full day of work over a period of 3 months [3]. LBP events

also lead to a significant financial burden for both individuals and clinical facilities, with combined direct and indirect costs of treatment for musculoskeletal injuries and associated pain estimated to be approximately US $213 billion annually [4].

LBP events fall into 2 major categories: acute and chronic [5]. Acute LBP occurs suddenly, usually associated with trauma or injury with subsequent pain, whereas chronic LBP is often reported by patients in regular checkups and has led to a significant increase in the use of health care services over the past two decades. It is very important to differentiate between acute and chronic LBP in the clinical setting as these conditions—as well as their management and billing—are substantively different. Chronic back pain is generally treated with spinal injections [6,7], surgery [8,9], and/or pain medications [10,11], whereas anti-inflammatories and a rapid return to normal activities of daily living are generally the best recommendations for acute LBP [12].

However, acute and chronic LBP are usually not explicitly separated in electronic health records (EHRs) because of a lack of distinguishing codes. The International Classification of Diseases, 10th revision (ICD-10) standard only includes the code M54.5 to characterize *low back pain* diagnosis, and it does not provide modifiers to distinguish different LBP acuities [13]. Acuity is usually reported in clinical notes, requiring a retrospective chart review of the free text to characterize LBP events, which is time consuming and not scalable [14]. Moreover, acuity can be expressed in different ways. For example, the text could mention *acute low back pain* or *acute LBP*, but could also simply report *shooting pain down into the lower extremities*, *limited spine range of motion*, *vertebral tenderness*, *diffuse pain in lumbar muscles*, and so on [15]. This variability makes it difficult for clinical facilities and researchers to group LBP episodes by acuity to perform key tasks, such as defining appropriate diagnostic and billing codes; evaluating the effectiveness of prescribed treatments; and deriving data-driven therapeutic guidelines and improved diagnostic methods that could reduce time, disability, and cost.

This paper is the first to explore the use of automated approaches based on machine learning and information retrieval to analyze free-text clinical notes and identify the acuity of LBP episodes. Specifically, we use a set of manually annotated notes to train and evaluate various machine learning architectures based on logistic regression (LR), n-grams, topic models, word embeddings, and convolutional neural networks, and to demonstrate that some of these models are able to identify acute LBP episodes with promising precision. In addition, we demonstrate the ineffectiveness of using ICD-10 codes alone to train the models, reinforcing the idea that they are not sufficient to differentiate the acuity of LBP. Our overall objective was to build an automated framework that can help front line primary care providers (PCPs) in the development of targeted strategies and return-to-work (RTW) options for acute LBP episodes in clinical practice.

Background and Significance

PCPs are commonly the first medical practitioners to assess patient's musculoskeletal injuries and pain associated with these injuries and are, therefore, in a unique position to offer reassurance, treatment options, and RTW recommendations catered to the acuity of the injury and pain associated with it. Several studies have documented increases in medication prescriptions and visits to physicians, physical therapists, and chiropractors for LBP episodes [16-18]. As individuals with chronic LBP seek care and use health care services more frequently than those with acute LBP, increases in health care use and costs for back pain are driven more by chronic than acute cases [19].

A rapid return to normal activities of daily living, including work, is generally the best activity recommendation for acute LBP management [12]. The number of workdays that are lost because of acute LBP can be reduced by implementing clinical practice guidelines in the primary care setting [20]. In previous work, Cruz et al built an RTW protocol tool for PCPs based on guidelines from the LBP literature [21]. On the basis of the type of work (eg, clerical, manual, or heavy) and the severity of the condition, the doctor would recommend RTW options (in partial or full duty capacity) within a certain number of days. The study found that physicians were likely to use this protocol, especially when it was integrated into the EHRs. However, the protocol was not always used for patients suffering from acute LBP as the research team was unable to quickly identify the acuity using only the structured EHR data (eg., ICD-10 codes). Acuity information was only available in the progress notes and was thus not incorporated into the automated recommendations. This prevented the research team from providing accurate feedback to PCPs based on a full picture of the patient's condition. A similar tool that could incorporate acuity information from notes could provide much more specific recommendations to PCPs that incorporate best practice guidelines for each acuity level. Besides leading to more precise care, this would streamline billing for LBP [22]. Similar needs arise for other musculoskeletal conditions, such as knee, elbow, and shoulder pain, where ICD-10 codes do not differentiate by pain level and acuity [23,24].

Machine learning methods for EHR data processing are enabling improved understanding of patient clinical trajectories, creating opportunities to derive new clinical insights [25,26]. In recent years, the application of deep learning, a hierarchical computational design based on layers of neural networks [27], to structured EHRs has led to promising results on clinical tasks such as disease phenotyping and prediction [28-33]. However, a wealth of relevant clinical information remains locked behind clinical narratives in the free text of notes. Natural language processing (NLP)—a branch of computer science that enables machines to process human language [34] for applications such as machine translation [35], text generation [36], and image captioning [37]—has been used to parse clinical notes to extract relevant insights that can guide clinical decisions [38]. Recent applications of deep learning to clinical NLP have classified clinical notes according to diagnosis or disease codes [39-41], predicted disease onset [32,42], and extracted primary cancer sites and their laterality in pathology reports [43,44]. However, although deep learning has successfully been applied to analyze clinical notes, traditional methods are still preferable when training data are limited [45,46].

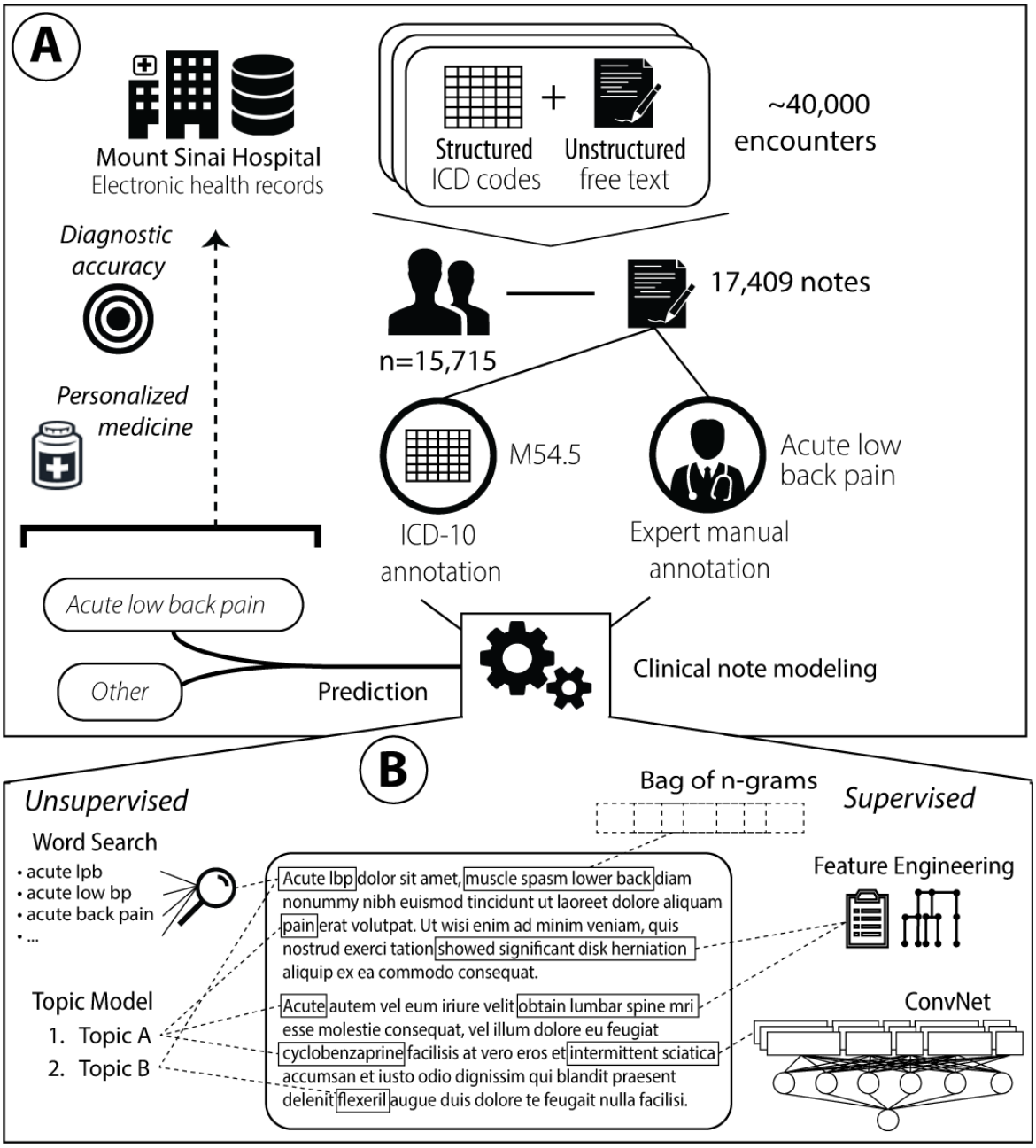
Regardless of the specific methodology, tools based on NLP applied to clinical narratives have not been widely used in clinical settings [31,38], despite the fact that physicians are likely to follow computer-assisted guidelines if recommendations are tied to their own observations [47]. In this paper, we present an NLP-based framework that can help physicians adhere to best practices and RTW recommendations for LBP. To the best of our knowledge, there are no studies to date that have applied machine learning to clinical notes to distinguish the acuity of a musculoskeletal condition in cases where it is not explicitly coded.

Methods

Overview

The conceptual steps of this study are summarized in Figure 1, specifically dataset composition, text processing, clinical notes modeling, and experimental evaluation. The overall goal was to evaluate the feasibility of automatically identifying clinical notes reporting *acute LBP* episodes.

Figure 1. Conceptual framework used to evaluate the use of automated approaches based on machine learning and information retrieval to analyze free-text clinical notes and identify acute low back pain episodes (a). The various unsupervised and supervised machine learning approaches used for clinical note modeling (b). ConvNet: convolutional neural network-based architecture; ICD-10: international classification of diseases, 10th revision.



Dataset

We used a set of free-text clinical notes extracted from the Mount Sinai data warehouse, made available for use under institutional review board approval following Health Insurance Portability and Accountability Act guidelines. The Mount Sinai Health System is an urban tertiary care hospital located in the Upper East Side of Manhattan in New York City. It generates

a high volume of structured, semistructured, and unstructured data as part of its routine health care and clinical operations, which include inpatient, outpatient, and emergency room visits. These clinical notes were collected during a previous pilot study evaluating an RTW tool based on EHR data that included nearly 40,000 encounters for 15,715 patients spanning from 2016 to 2018 and clinical notes written by 81 different providers [21]. In that study, we used the published literature to develop a list of guidelines to determine the assessment and management of acute LBP episodes in clinical practice. In particular, we used ICD-10 codes and other parameters, such as *presenting complaint*, *pre-existing conditions*, *management factors*, and *imaging/radiology/test ordered*, to define and label the acuity of LBP in a clinical encounter. Following these guidelines, 14 individuals (physical medicine and rehabilitation fellows, residents, and medical students) manually reviewed a random set of 4291 clinical notes associated with these encounters and labeled all *acute low back pain* events. Each note was reviewed by at least two individuals and was further checked by a lead physician researcher if it was marked as ambiguous and/or there was discordance between reviewers.

This project leveraged the entire set of clinical notes that were collected in the previous study. In particular, we joined all the progress notes of these encounters under the same initial visit, and we eliminated duplicate, short (less than 3 words), and nonmeaningful reports. The final dataset was composed of 17,409 distinct clinical notes, with length ranging from 7 to 6638 words. Of this set, 3092 notes were manually reviewed in the previous study, and 891 of them were annotated as *acute LBP*. The remaining 14,317 notes were not manually evaluated and were related to different clinical domains, including various musculoskeletal disorders and potentially LBP events. In this final dataset, 1973 notes were also associated with an encounter billed with an ICD-10 M54.5 *Low back pain* code.

Text Processing

Every note in the dataset was tokenized, divided into sentences, and checked to remove punctuation; numbers; and nonrelevant concepts such as URLs, emails, and dates. Each note was then represented as a list of sentences, with every sentence being a list of lemmatized words represented as one-hot encodings. The vocabulary was composed of all the words appearing at least five times in the training set. The discarded words were corrected to the terms in the vocabulary having the minimum edit distance, that is, the minimum number of operations required to transform one string into the other [48]. This step reduced the number of misspelled words and prevented the accidental discarding of relevant information; at the same time, it also limited the size of the vocabulary to improve scalability [39]. Overall, the vocabulary covering the whole dataset comprised 56,142 unique words.

Clinical Note Modeling

We evaluated different approaches for identifying clinical notes that refer to acute LBP episodes. These included both supervised and unsupervised methods. Although we benefited from the use of high-quality manual annotations to train the supervised models, we also investigated alternatives that did not require manual annotation of notes. All these methods provided straightforward explanations of their predictions, enabling us to validate each model and to identify parts of text and patterns that are relevant to the *acute LBP* predictions.

Keyword Search

We searched for a set of relevant keywords in the text. In particular, we looked for “acute low back pain,” “acute lbp,” “acute low bp,” and “acute back pain,” and we counted their occurrences in the text. We used the NegEx algorithm [49] to annotate and remove negated occurrences of the keywords. In the evaluation, we refer to this model as *WordSearch*.

Topic Modeling

We used topic modeling on the full set of words contained in the notes to capture abstract topics referred to in the dataset [50]. Topic modeling is an unsupervised inference process, in this case, implemented using latent Dirichlet allocation [51], which captures patterns of word co-occurrences within documents to define interpretable topics (ie, multinomial distribution of words) and represent a document as a multinomial over these topics. Every document can then be classified as talking about 1 or (usually) more topics. Topic modeling is often used in health care to generalize clinical notes, improve the automatic processing of patient data, and explore clinical datasets [52-55].

In this study, we assumed that 1 or more of these topics might refer to acute LBP. To discover them, we identified the most likely topics for a set of keywords (ie, “acute,” “low,” “back,” “pain,” “lbp,” and “bp”), and we manually reviewed them to retain only those that seemed more likely to characterize acute LBP episodes (ie, that included most of the keywords with high probability). We then considered the maximum likelihood among these topics as the probability that a report referred to acute LBP (ie, *TopicModel* in the experiments).

Bag of N-Grams

Each clinical note was represented as a bag of n-grams (BoN; with n=1, ..., 5), with term frequency-inverse document frequency (TF-IDF) weights (determined from the corpus of documents). Each n-gram is a contiguous sequence of *n* words from the text. We considered all the words in the vocabulary and filtered the common stop words based on the English dictionary before building all the n-grams. The classification was implemented using LR with least absolute shrinkage and selection operator (LASSO; ie, *BoN-LR*).

Feature Engineering

We used the protocol built by Cruz et al [21] to define acute LBP episodes in the clinical notes. In particular, we used all the concepts described in that guideline, preprocessed them with the same algorithm used for the clinical notes, and built a set of 5154 distinct n-grams (with n=1, ..., 5), which we refer to as *FeatEng*. We then represented each clinical note as a bag of FeatEng (ie, we counted the occurrences of only these n-grams in the text), normalized with TF-IDF weights, and classified them using LR with LASSO (ie, *FeatEng-LR*).

Deep Learning

We implemented an end-to-end deep neural network architecture based on convolutional neural networks that takes as input the full note and outputs its probability of being related to *acute LBP* (ie, *ConvNet* in the experiments). The first layer of the architecture maps the words to dense vector representations (ie, *embeddings*), which attempt to contextualize the semantic meaning of each word by creating a metric space where vectors of semantically similar words are close to each other. We applied word2vec with the skip-gram algorithm to the parsed notes [56] to initialize the embedding of each word in the vocabulary. Word2vec is commonly used with EHRs to learn embeddings of medical concepts from structured data and clinical notes [46,57-59].

The embeddings were then fed to a convolutional neural network inspired by the model described by Kim [60] and Liu et al [42]. This architecture concatenates representations of the text at different levels of abstraction by essentially choosing the most relevant n-grams at each level. Here, we first applied a set of parallel 1 dimensional (1D) convolutions on the input sequence with kernel sizes ranging from 1 to 5, thus simulating n-grams with n=1, ..., 5. The outputs of each of these convolutions were then max-pooled over the whole sequence and concatenated to a $5 \times d$ dimensional vector, where d is the number of 1D convolutional filters. This representation was then fed to sequences of fully connected layers, which learn the interactions between the text features, and finally to a sigmoid layer that outputs the prediction probability.

The n-grams that are most relevant to the prediction, in this architecture, are those that activate the neurons in the max-pooling layer. Therefore, we used the log-odds that the n-gram contributes to the sigmoid decision function [42] as an indication of how much each n-gram influences the decision.

Evaluation Design

We evaluated all the architectures using a 10-fold cross-validation experiment, with every note appearing in the test set only once. In each training set, we used a random 90/10 split to train and validate all the model configurations. As baseline, we also report the results obtained by considering as *acute LBP* all the notes associated with the *Low back pain* M54.5 ICD-10 code (ie, *ICD-10* in the results).

Training Annotations

We considered 2 different sets of annotations as gold standards to train the supervised models. In the first experiment, we used the manually curated annotations provided with the dataset from previous work [21], whereas in the second experiment, we trained the models using the ICD-10 codes associated with each note encounter. Both experiments were evaluated using manual annotations. The rationale was to compare the feasibility of identifying acute LBP events when manual annotations are and are not available. We trained the classifier to output *acute LBP* versus *other* because the goal of the project was to identify clinical notes with acute LBP events rather than discriminate different facets of LBP events (eg, *chronic LBP* vs *acute LBP*).

Metrics

For all experiments, we report area under the receiver operating characteristic curve (AUC-ROC); precision, recall, and F score; and area under the precision-recall curve (AUC-PRC) [61]. The ROC curve is a plot of true positive rate versus false positive rate found over the set of predictions. F score is the harmonic mean of classification precision and recall per annotation, where precision is the number of correct positive results divided by the number of all positive results, and recall is the number of correct positive results divided by the number of positive results that should have been returned. The PRC is a plot of precision and recall for different thresholds. The areas under the ROC and PR curves are computed by integrating the corresponding curves.

Model Hyperparameters

The model hyperparameters were empirically tuned using the validation sets to optimize the results with both training annotations. In the topic modeling method, we inferred topics using the whole training set of documents and 200 topics (derived using perplexity analysis). Although seemingly more intuitive, using only the notes associated with the M54.5 *Low back pain* ICD-10 code actually produced worse results. For each fold, the most relevant topics associated with acute LBP were manually reviewed and used to annotate the notes. In the deep learning architecture, we used embeddings with size 300 and full-length notes. We trained word2vec just on the clinical note dataset to initialize embeddings. Preinitializing the embeddings with a general-purpose corpus did not lead to any improvement. Each convolutional neural network had 200 filters and used a rectified linear unit (ReLu) activation function. We added 2 fully connected layers of size 600 following the convolutional neural networks with ReLu activations and batch normalization. Dropout values across the layers were all set to 0.5. The architecture was trained using cross-entropy loss with the Adam optimizer for 5 epochs and batch size 32 (learning rate=0.001). The classification thresholds for precision, recall, and F score were found by ranging the value from 0.1 to 1, with 0.1 increments, and retaining, for each model, the value leading to the best results on the validation set.

Results

Table 1 and Figure 2 show the average results of the 10-fold cross-validation experiment for all the models considered. The best results were obtained by convolutional neural network-based architecture (ConvNet) when trained with the manual annotations. Although this is not entirely surprising given the success of deep learning for NLP when high-quality annotations and a large amount of data (ie, on the order of millions of training examples) are available, this was not certain in this domain where the training dataset was much smaller. As expected, the results obtained by the baseline and by training the models using the ICD-10 codes were not as good, confirming that the M54.5 ICD-10 code is not a sufficient indicator of acute LBP. TopicModel leads to similar performance but provides a more intuitive and potentially effective way for exploring the collection, extracting meaningful patterns that are related to acute LBP episodes. The most relevant topics included words

defining acute LBP (eg, acute, low, back, pain, lbp, spasm, lifting, sciatica) and also included several medications that are usually prescribed to treat inflammation and pain (eg, Cyclobenzaprine, Flexeril, and Advil). Although this approach might not be robust enough for clinical application, a refined and manually curated version of TopicModel promises to allow an efficient prefiltering of clinical reports that can speed up the manual work required to annotate them. On the contrary, but as expected, WordSearch performed poorly as the condition is mentioned in too many different ways across the text, and simple keywords were not sufficient.

Table 1. The classification results in identifying clinical notes with acute low back pain (LBP) episodes averaged over the 10-fold cross-validation experiment. We compared different supervised and unsupervised strategies: keyword search (WordSearch), topic modeling (TopicModel), logistic regression with bag of n-grams (BoN-LR) and manual features (FeatEng-LR), and deep learning (ConvNet). The supervised models (ie, BoN-LR, FeatEng-LR, and ConvNet) were trained using manual annotations or M54.5 International Classification of Diseases, 10th revision (ICD-10) codes. The ICD-10 baseline simply considered as acute LBP all the notes associated with the generic M54.5 Low back pain ICD-10 code.

Model	Precision	Recall	F score	Area under the receiver operating characteristic curve	Area under the precision-recall curve
Baseline					
ICD-10[a]	0.32	0.68	0.41	0.81	0.42
Unsupervised methods					
WordSearch	0.71	0.03	0.06	0.52	0.40
TopicModel	0.44	0.58	0.50	0.92	0.46
Trained with the M54.5 ICD-10 code					
BoN-LR[b]	0.50	0.70	0.59	0.83	0.42
FeatEng-LR[c]	0.47	0.59	0.52	0.88	0.41
ConvNet[d]	0.55	0.68	0.61	0.89	0.46
Trained with manual annotations					
BoN-LR	0.53	0.64	0.58	0.93	0.56
FeatEng-LR	0.58	0.66	0.62	0.93	0.58
ConvNet	0.65	0.73	0.70	0.98	0.72

[a]ICD-10: International Classification of Diseases, 10th revision codes.

[b]BoN-LR: logistic regression with bag of n-grams.

[c]FeatEng-LR: logistic regression with feature engineering.

[d]ConvNet: convolutional neural network-based architecture.

Figure 2. Receiver operating characteristic and precision-recall curves obtained when using as training data for BoN-LR, FeatEng-LR and ConvNet the manual annotations (a) and the M54.5 ICD-10 codes (b). ConvNet trained using the manual annotations obtained the best results. In the absence of manual annotations to use for training, TopicModel worked better than methods trained using ICD-10 codes, which proved not to be a good indicator to identify acuity in low back pain episodes. BoN-LR: logistic regression with bag of n-grams; ConvNet: convolutional neural network-based architecture; FeatEng-LR: logistic regression with feature engineering; ICD-10: international classification of diseases, 10th revision; PR: precision-recall; ROC: receiver operating characteristic.

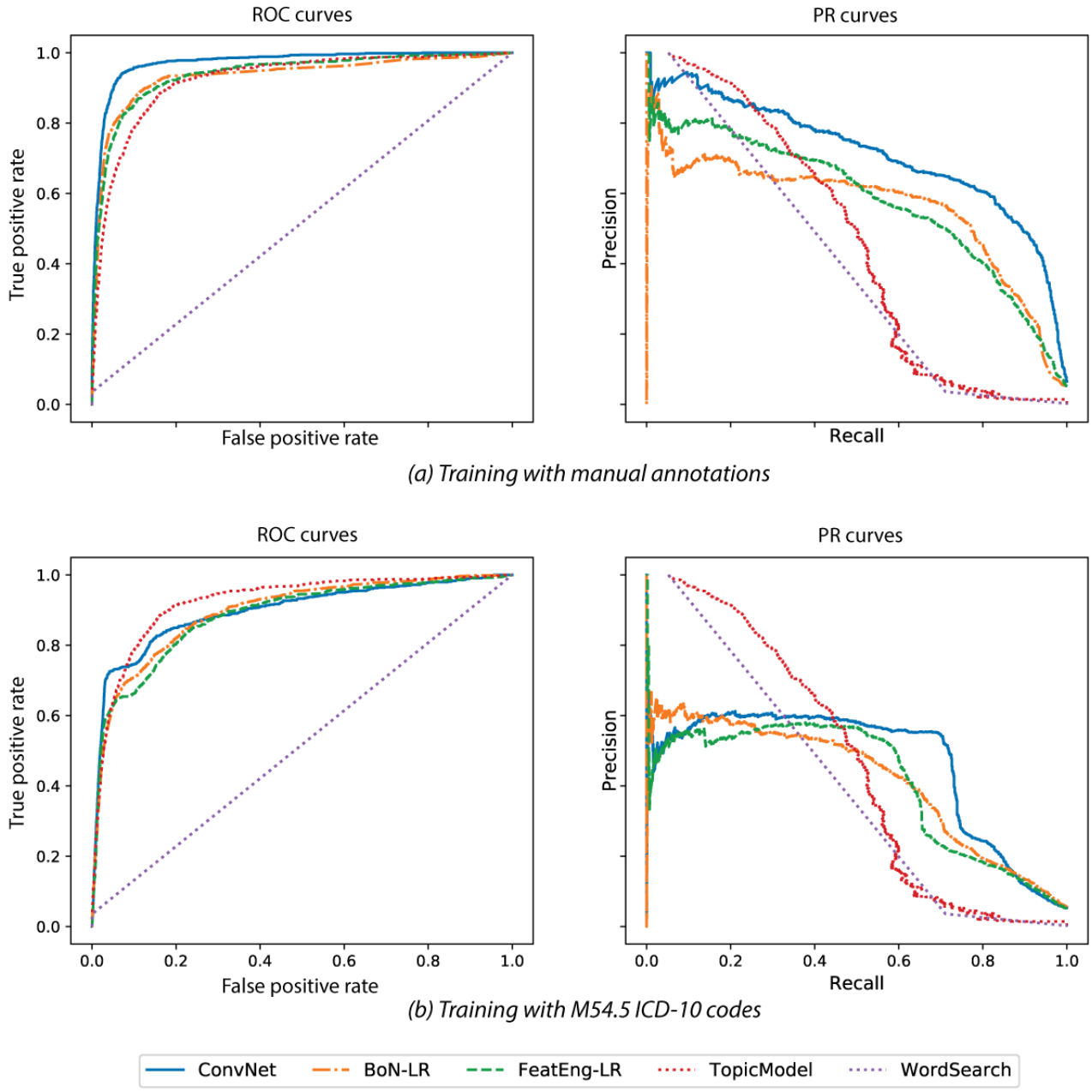


Figure 3 shows the classification results in terms of AUC-ROC and AUC-PRC when randomly subsampling the *acute LBP* manual annotations in the training set. We found that ConvNet always outperforms the other methods based on LR as well as TopicModel. In addition, we notice that using just 240 out of 800 (30.0%) manual annotations in the training set already leads to better results than using ICD-10 codes as training labels. This is a particularly interesting insight as it shows that only minimal manual work is required to achieve good classifications; these can then be further improved by adding automatically annotated notes to the model (after manual verification) and retraining.

Figure 3. Area under the receiver operating characteristic and precision-recall curves obtained when training the supervised models using random subsamples of the manual annotations. TopicModel is reported as reference baseline. ConvNet obtained satisfactory results when trained using less manually annotated documents, showing robustness and scalability to the gold standard. AUC-PRC: area under the precision-recall curve; AUC-ROC: area under the receiver operating characteristic curve; BoN-LR: logistic regression with bag of n-grams; ConvNet: convolutional neural network-based architecture; FeatEng-LR: logistic regression with feature engineering.

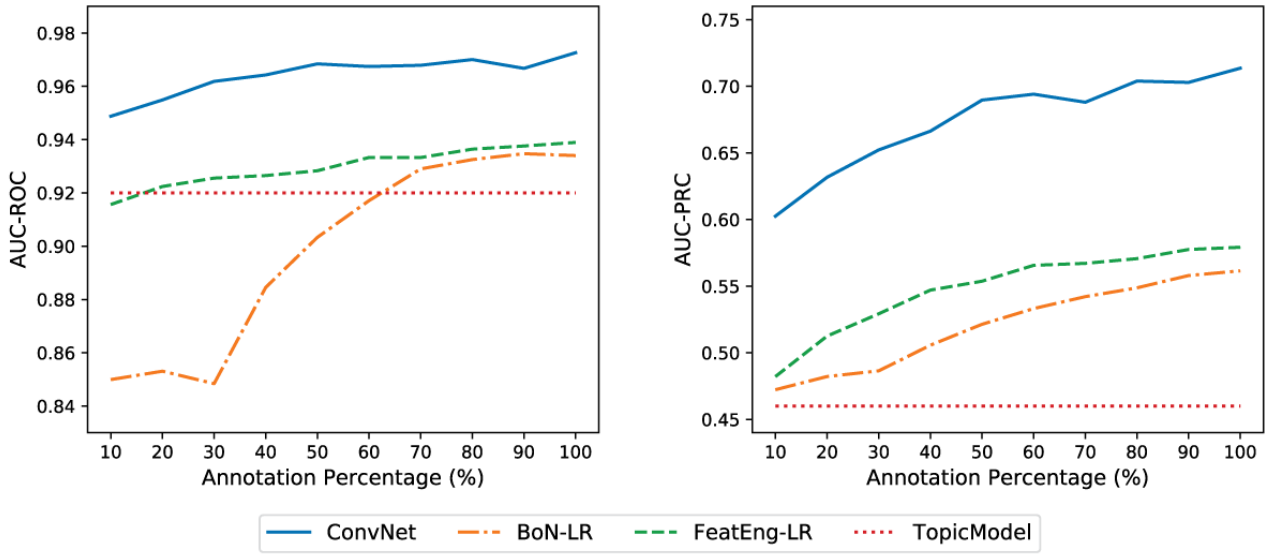
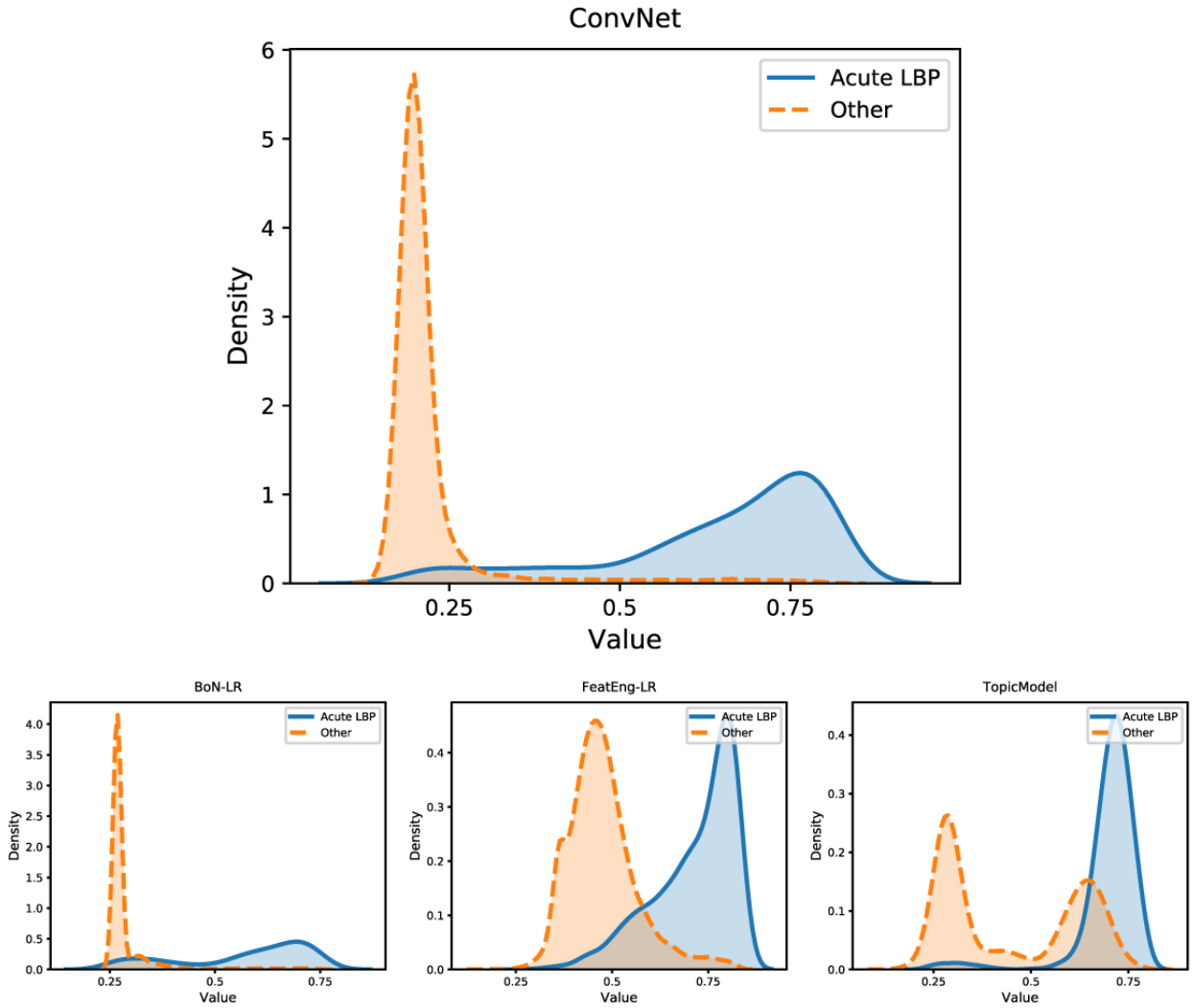


Figure 4 highlights the distributions of the classification scores (predicted probability of the label *acute LBP*) derived by several supervised models (trained with manual annotations) and TopicModel. ConvNet shows a clear separation between acute LBP notes and the rest of the dataset. In particular, all acute LBP notes had scores greater than 0.2, with 81.6% (727/891) of them having scores greater than 0.5. On the contrary, only 347 controls had scores greater than 0.5, meaning that only a few notes were highly likely to be misclassified. Similarly, TopicModel had no controls with scores greater than 0.7, and all acute LBP notes had scores greater than 0.2.

Figure 4. Representation of the probability distribution of the scores obtained by BoN-LR, FeatEng-LR, ConvNet, and TopicModel. ConvNet led to a good separation between acute low back pain clinical notes and all the other documents. In other cases, such separation is not as clear, explaining the worse classification results obtained by those models. BoN-LR: logistic regression with bag of n-grams; ConvNet: convolutional neural network-based architecture; FeatEng-LR: logistic regression with feature engineering.



Finally, Table 2 summarizes some of the n-grams driving the *acute LBP* predictions obtained by ConvNet (trained with manual annotations) across the experiments. Although some of these are obvious and refer to the disease itself (eg, “acute lbp”), others refer to medications (eg, “prescribed muscle relaxant” and “flexeril”) and recommendations (eg, “rtw full duty quick”). Given their clinical meaning and relevance, all these patterns can be further analyzed and reviewed to potentially drive the development of guidelines for, for example, treatment and RTW options.

Table 2. Examples of n-grams that were relevant in identifying acute low back pain notes when using convolutional neural network-based architecture trained with manual annotations. The n-grams' relevance was determined by analyzing the neurons of the convolutional neural networks activating the max-pooling layers and their log-odds to contribute to the final output. Log-odds were filtered per notes and then averaged over all the notes and evaluation folds.

Type	Acute LBP[a]-related predictive n-grams
Diagnosis	• Muscle spasm lower back • Acute LBP flare • Been having acute back pain • Acute midline LBP • Sports acute bilateral LBP • Acute low back pain • Acute LBP
Related conditions	• Gait abnormality • Showed significant disk herniation • Intermittent sciatica • Spinal stenosis
Medications	• Back pain flare prescribed flexeril • Cyclobenzaprine • Flexeril • Naproxen for acute low back • Prescribed muscle relaxant
Recommendations	• Back brace for back pain • Obtain lumbar spine MRI[b] • Recommendation RTW[c] visit • RTW full duty quick

[a]LBP: low back pain.

[b]MRI: magnetic resonance imaging.

[c]RTW: return-to-work.

Discussion

Principal Findings

In this work, we evaluated the use of several machine learning approaches to identify acute LBP episodes in free-text clinical notes to better personalize the treatment and management of this condition in primary care. The experimental results showed that it is possible to extract acute LBP episodes with promising precision, especially when at least some manually curated annotations are available. In this scenario, ConvNet, a deep learning architecture based on convolutional neural networks, significantly outperformed other shallow techniques based on BoN and LR, opening the possibility to boost performances using more complex architectures from current research in the NLP community. The implemented deep architecture also provides an easy mechanism to explain the predictions, leading to informed decision support based on model transparency [62,63] and the identification of meaningful patterns that can drive clinical decision making. If no annotations are available, experiments showed that the use of topic modeling is preferred to training a classifier using only the M54.5 ICD-10 codes (ie, *Low back pain*) associated with the clinical note encounter, which proved to be a poor indicator to discriminate LBP episodes. In addition, the topics identified can serve as an intuitive tool to inform guidelines and recommendations, to prefilter the documents, and to reduce the manual work required to annotate the notes. The proposed framework is inherently domain agnostic and does not require any manual supervision to identify relevant features from the free text. Therefore, it can be leveraged in other musculoskeletal condition domains where acuity is not expressed in the ICD-10 diagnostic codes, such as knee, elbow, and shoulder pain.

Potential Applications

Medical care decisions are often based on heuristics and manually derived rule-based models constructed on previous knowledge and expertise [64]. Cognitive biases and personality traits, such as aversion to risk or ambiguity, overconfidence, and the anchoring effect, may lead to diagnostic inaccuracies and medical errors, resulting in mismanagement or inadequate utilization of resources [65]. In the LBP domain, this may lead to delays in finding the right therapy and assisting patients in the return to normal activities, increased risk of transitioning the condition from acute to chronic, discomfort for patients, and increased economic burdens on clinical facilities to adequately treat and manage this patient population. Deriving data-driven guidelines for treatment recommendations can help in reducing these cognitive biases and personality traits, leading to more consistent and accurate decisions. In this scenario, the proposed frameworks integrate seamlessly with the RTW tool proposed by Cruz et al [21] by including acuity-relevant information in the clinical notes and addressing 1 of the limitations of that study (ie, recommending the RTW tool at the point of care by accurately identifying the condition as acute LBP). Similarly, an understanding of the patterns driving the

predictions can lead to the development of new and improved treatment strategies for various types of injuries, which can be presented to the clinicians at the time of patient encounter to help them with better management of the condition. Although physicians will continue to have autonomy in determining optimal care pathways for their patients, the recommendations provided by the supporting framework will be useful to systematize and support their activities within the realm of the busy clinical practice. Posterior analysis of the clinical notes to infer acute LBP episodes can also help in assigning the proper diagnostic and billing codes for the encounter. In a foreseeable future scenario where, clinical observations are automatically transcribed via voice and EHRs are processed in real time, an automated tool that identifies acuity information could also improve the accuracy of diagnosis and billing in real time, with no need to wait for posterior evaluations.

Limitations

This work evaluated the feasibility of using machine learning to identify acute LBP episodes in clinical notes. Therefore, we compared different types of models (shallow vs deep) and learning frameworks (unsupervised vs supervised) to identify the best directions for implementation and deployment in real clinical settings. Although several of the architectures evaluated in this work obtained promising results, more sophisticated models are likely to improve these performances, especially in the deep learning domain. For example, algorithms based on attention models [66], Bidirectional Encoder Representations from Transformers [67], or XLNet [68] have shown encouraging results on similar NLP tasks and are likely to obtain better results in this domain as well. In this work, we only focused on processing clinical notes; however, embedding structured EHR data, especially medications, imaging studies, and/or laboratory tests, into the method should improve the results.

The dataset of clinical notes used in this study originated from a geographically diverse set of primary care clinics serving the New York City population across the city's metro area over a limited period (ie, 2016 to 2018). Providers were enrolled and randomized into the study on a rolling basis, with the number of encounters for LBP varying for each individual provider, based on his/her own practice. The majority of the PCPs were assistant professors serving on the front lines. No specialists were included in the initial study, as the pilot project was only geared toward the PCPs. Consequently, the results of this study might not be applicable to specialty care practice.

Future Work

The classification of LBP episodes as acute or chronic at the point of care level within primary care practice is imperative for an RTW tool to be effectively used to render evidence-based guidelines. At this time, we plan to classify a large set of notes, derive patterns related to acute LBP, and extend the tool proposed by Cruz et al [21] according to them. We further plan to identify cases where the RTW tool can be easily deployed based on EHR integration in the clinical domain. We will also begin to address some of the methodological limitations of this study to optimize performance and evaluate its generalizability outside primary care. Finally, we aim to evaluate the feasibility of this type of approach for other musculoskeletal conditions, in particular, shoulder and knee pain.

Conclusions

This study demonstrates the feasibility of using machine learning to automatically identify acute LBP episodes from clinical reports using only unstructured free-text data. In particular, manually annotating a set of notes to use as a gold standard can lead to effective results, especially when using deep learning. Topic modeling can help in speeding up the annotation process, initiating an iterative process where initial predictions are validated and then used to refine and optimize the model. This approach provides a generalizable framework for learning to differentiate disease acuity in primary care, which can more accurately and specifically guide the diagnosis and treatment of LBP. It also provides a clear path toward improving the accuracy of coding and billing of clinical encounters for LBP.

Acknowledgments

IN and LC would like to thank the Pilot Projects Research Training Program of the New York and New Jersey Education and Research Center, National Institute for Occupational Safety and Health, for their funding (grant #T42 OH 008422). RM is grateful for the support from the Hasso Plattner Foundation and a courtesy GPU donation from NVIDIA.

Authors' Contributions

RM and IN initiated the idea and wrote the manuscript. IN collected the data and provided clinical support. RM conducted the research and the experimental evaluation. BP advised on evaluation strategies and refined the manuscript. BG, HL, and LC refined the manuscript. JD supported the research. All the authors edited and reviewed the manuscript.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control Prevention (CDC). Prevalence and most common causes of disability among adults--United States, 2005. MMWR Morb Mortal Wkly Rep 2009 May 1;58(16):421-426 [FREE Full text] [Medline: 19407734]

2. Ricci JA, Stewart WF, Chee E, Leotta C, Foley K, Hochberg MC. Back pain exacerbations and lost productive time costs in United States workers. Spine (Phila Pa 1976) 2006 Dec 15;31(26):3052-3060. [doi: 10.1097/01.brs.0000249521.61813.aa] [Medline: 17173003]
3. Luckhaupt SE, Dahlhamer JM, Gonzales GT, Lu ML, Groenewold M, Sweeney MH, et al. Prevalence, recognition of work-relatedness, and effect on work of low back pain among US workers. Ann Intern Med 2019 May 14. [doi: 10.7326/M18-3602] [Medline: 31083729]
4. BMUS: The Burden of Musculoskeletal Diseases in the United States. Health Care Utilization and Economic Cost URL: https://www.boneandjointburden.org/2014-report/if0/health-care-utilization-and-economic-cost [accessed 2019-04-22]
5. Fairbank J, Gwilym SE, France JC, Daffner SD, Dettori J, Hermsmeyer J, et al. The role of classification of chronic low back pain. Spine (Phila Pa 1976) 2011 Oct 1;36(21 Suppl):S19-S42. [doi: 10.1097/BRS.0b013e31822ef72c] [Medline: 21952188]
6. Weiner DK, Kim YS, Bonino P, Wang T. Low back pain in older adults: are we utilizing healthcare resources wisely? Pain Med 2006;7(2):143-150. [doi: 10.1111/j.1526-4637.2006.00112.x] [Medline: 16634727]
7. Friedly J, Chan L, Deyo R. Increases in lumbosacral injections in the Medicare population: 1994 to 2001. Spine (Phila Pa 1976) 2007 Jul 15;32(16):1754-1760. [doi: 10.1097/BRS.0b013e3180b9f96e] [Medline: 17632396]
8. Deyo RA, Mirza SK. Trends and variations in the use of spine surgery. Clin Orthop Relat Res 2006 Feb;443:139-146. [doi: 10.1097/01.blo.0000198726.62514.75] [Medline: 16462438]
9. Deyo RA, Nachemson A, Mirza SK. Spinal-fusion surgery - the case for restraint. N Engl J Med 2004 Feb 12;350(7):722-726. [doi: 10.1056/NEJMsb031771] [Medline: 14960750]
10. Ballantyne JC. Opioids for the treatment of chronic pain: mistakes made, lessons learned, and future directions. Anesth Analg 2017 Nov;125(5):1769-1778. [doi: 10.1213/ANE.0000000000002500] [Medline: 29049121]
11. Luo X, Pietrobon R, Hey L. Patterns and trends in opioid use among individuals with back pain in the United States. Spine (Phila Pa 1976) 2004 Apr 15;29(8):884-90; discussion 891. [doi: 10.1097/00007632-200404150-00012] [Medline: 15082989]
12. Malmivaara A, Häkkinen U, Aro T, Heinrichs ML, Koskenniemi L, Kuosma E, et al. The treatment of acute low back pain--bed rest, exercises, or ordinary activity? N Engl J Med 1995 Feb 9;332(6):351-355. [doi: 10.1056/NEJM199502093320602] [Medline: 7823996]
13. ICD-10 Data. 2020 ICD-10-CM Diagnosis Code M54.5: Low back pain URL: https://www.icd10data.com/ICD10CM/Codes/M00-M99/M50-M54/M54-/M54.5 [accessed 2020-01-07]
14. Petersen T, Laslett M, Juhl C. Clinical classification in low back pain: best-evidence diagnostic rules based on systematic reviews. BMC Musculoskelet Disord 2017 May 12;18(1):188 [FREE Full text] [doi: 10.1186/s12891-017-1549-6] [Medline: 28499364]
15. Casazza BA. Diagnosis and treatment of acute low back pain. Am Fam Physician 2012 Feb 15;85(4):343-350 [FREE Full text] [Medline: 22335313]
16. Feuerstein M, Marcus SC, Huang GD. National trends in nonoperative care for nonspecific back pain. Spine J 2004;4(1):56-63. [doi: 10.1016/j.spinee.2003.08.003] [Medline: 14749194]
17. Kessler RC, Davis RB, Foster DF, van Rompay MI, Walters EE, Wilkey SA, et al. Long-term trends in the use of complementary and alternative medical therapies in the United States. Ann Intern Med 2001 Aug 21;135(4):262-268. [doi: 10.7326/0003-4819-135-4-200108210-00011] [Medline: 11511141]
18. Martin BI, Deyo RA, Mirza SK, Turner JA, Comstock BA, Hollingworth W, et al. Expenditures and health status among adults with back and neck problems. J Am Med Assoc 2008 Feb 13;299(6):656-664. [doi: 10.1001/jama.299.6.656] [Medline: 18270354]
19. Freburger JK, Holmes GM, Agans RP, Jackman AM, Darter JD, Wallace AS, et al. The rising prevalence of chronic low back pain. Arch Intern Med 2009 Feb 9;169(3):251-258 [FREE Full text] [doi: 10.1001/archinternmed.2008.543] [Medline: 19204216]
20. Rossignol M, Abenhaim L, Séguin P, Neveu A, Collet JP, Ducruet T, et al. Coordination of primary health care for back pain. A randomized controlled trial. Spine (Phila Pa 1976) 2000 Jan 15;25(2):251-8; discussion 258. [doi: 10.1097/00007632-200001150-00018] [Medline: 10685491]
21. Cruz LC, Alamgir HA, Sheth P, Nabeel I. Development of a return to work tool for primary care providers for patients with low back pain: A pilot study. J Family Med Prim Care 2018;7(6):1185-1192 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_262_18] [Medline: 30613495]
22. Owens JD, Hegmann KT, Thiese MS, Phillips AL. Impacts of adherence to evidence-based medicine guidelines for the management of acute low back pain on costs of worker's compensation claims. J Occup Environ Med 2019 Jun;61(6):445-452. [doi: 10.1097/JOM.0000000000001593] [Medline: 31167221]
23. Coding Strategies. Reporting Pain in ICD-10-CM URL: https://www.codingstrategies.com/news/reporting-pain-icd-10-cm [accessed 2020-01-07]
24. Gross DP, Armijo-Olivo S, Shaw WS, Williams-Whitt K, Shaw NT, Hartvigsen J, et al. Clinical decision support tools for selecting interventions for patients with disabling musculoskeletal disorders: a scoping review. J Occup Rehabil 2016 Sep;26(3):286-318 [FREE Full text] [doi: 10.1007/s10926-015-9614-1] [Medline: 26667939]

25. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 May 2;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]
26. Glicksberg BS, Johnson KW, Dudley JT. The next generation of precision medicine: observational studies, electronic health records, biobanks and continuous monitoring. Hum Mol Genet 2018 May 1;27(R1):R56-R62. [doi: 10.1093/hmg/ddy114] [Medline: 29659828]
27. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
28. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016 May 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
29. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2018 Nov 27;19(6):1236-1246 [FREE Full text] [doi: 10.1093/bib/bbx044] [Medline: 28481991]
30. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: predicting clinical events via recurrent neural networks. JMLR Workshop Conf Proc 2016 Aug;56:301-318 [FREE Full text] [Medline: 28286600]
31. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc 2018 Oct 1;25(10):1419-1428 [FREE Full text] [doi: 10.1093/jamia/ocy068] [Medline: 29893864]
32. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018;1:18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
33. Miotto R, Li L, Dudley JT. Deep learning to predict patient future diseases from the electronic health records. In: Ferro N, Crestani F, Moens MF, editors. Advances in Information Retrieval. Cham: Springer; 2016:768-774.
34. Goldberg Y. A primer on neural network models for Natural Language Processing. J Artif Intell Res 2016;57(7):345-420. [doi: 10.1613/jair.4992]
35. Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, et al. arXiv e-Print archive. 2016. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation URL: http://arxiv.org/abs/1609.08144 [accessed 2020-01-07]
36. Kannan A, Kurach K, Ravi S, Kaufmann T, Tomkins A, Miklos B, et al. Smart Reply: Automated Response Suggestion for Email. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Fransisco, CA: ACM; 2016 Presented at: KDD'16; August 13-17, 2016; San Francisco, United States p. 955-964.
37. Vinyals O, Toshev A, Bengio S, Erhan D. Show and Tell: A Neural Image Caption Generator. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. 2015 Presented at: CVPR'15; June 7-12, 2015; Boston, MA URL: https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Vinyals_Show_and_Tell_2015_CVPR_paper.pdf
38. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural Language Processing of clinical notes on chronic diseases: systematic review. JMIR Med Inform 2019 Apr 27;7(2):e12239 [FREE Full text] [doi: 10.2196/12239] [Medline: 31066697]
39. Baumel T, Nassour-Kassis J, Cohen R, Elhadad M, Elhadad N. arXiv e-Print archive. 2017. Multi-Label Classification of Patient Notes a Case Study on ICD Code Assignment URL: http://arxiv.org/abs/1709.09587 [accessed 2020-01-07]
40. Mullenbach J, Wiegreffe S, Duke J, Sun J, Eisenstein J. arXiv e-Print archive. 2018. Explainable Prediction of Medical Codes from Clinical Text URL: http://arxiv.org/abs/1802.05695 [accessed 2020-01-07]
41. Shi H, Xie P, Hu Z, Zhang M, Xing EP. arXiv e-Print archive. 2017. Towards Automated ICD Coding Using Deep Learning URL: http://arxiv.org/abs/1711.04075 [accessed 2020-01-07]
42. Liu J, Zhang Z, Razavian N. arXiv e-Print archive. 2018. Deep EHR: Chronic Disease Prediction Using Medical Notes URL: http://arxiv.org/abs/1808.04928 [accessed 2020-01-07]
43. Yoon HJ, Ramanathan A, Tourassi G. Multi-task deep neural networks for automated extraction of primary site and laterality information from cancer pathology reports. In: Angelov P, Manolopoulos Y, Iliadis L, Roy A, Vellasco M, editors. Advances in Big Data. Cham, Switzerland: Springer; 2016:195-204.
44. Qiu JX, Yoon HJ, Fearn PA, Tourassi GD. Deep learning for automated extraction of primary sites from cancer pathology reports. IEEE J Biomed Health Inform 2018 Jan;22(1):244-251. [doi: 10.1109/JBHI.2017.2700722] [Medline: 28475069]
45. Turner CA, Jacobs AD, Marques CK, Oates JC, Kamen DL, Anderson PE, et al. Word2Vec inversion and traditional text classifiers for phenotyping lupus. BMC Med Inform Decis Mak 2017 Aug 22;17(1):126 [FREE Full text] [doi: 10.1186/s12911-017-0518-1] [Medline: 28830409]
46. Gehrmann S, Dernoncourt F, Li Y, Carlson ET, Wu JT, Welt J, et al. arXiv e-Print archive. 2017. Comparing Rule-Based and Deep Learning Models for Patient Phenotyping URL: http://arxiv.org/abs/1703.08705 [accessed 2020-01-07]
47. Davis DA, Taylor-Vaisey A. Translating guidelines into practice. A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. Can Med Assoc J 1997 Aug 15;157(4):408-416 [FREE Full text] [Medline: 9275952]
48. Navarro G. A guided tour to approximate string matching. ACM Comput Surv 2001;33(1):31-88 [FREE Full text] [doi: 10.1145/375360.375365]

49. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform 2001 Oct;34(5):301-310 [FREE Full text] [doi: 10.1006/jbin.2001.1029] [Medline: 12123149]
50. Blei DM. Probabilistic topic models. Commun ACM 2012;55(4):77. [doi: 10.1145/2133806.2133826]
51. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
52. Miotto R, Weng C. Case-based reasoning using electronic health records efficiently identifies eligible patients for clinical trials. J Am Med Inform Assoc 2015 Apr;22(e1):e141-e150 [FREE Full text] [doi: 10.1093/jamia/ocu050] [Medline: 25769682]
53. Perotte AJ, Wood F, Elhadad N, Bartlett N. Hierarchically Supervised Latent Dirichlet Allocation. In: Proceedings of the Advances in Neural Information Processing Systems 24. New York, NY: Curran Associates, Inc; 2011 Presented at: NIPS'11; December 12-17, 2011; Granada, Spain p. 2609-2617 URL: https://papers.nips.cc/paper/4313-hierarchically-supervised-latent-dirichlet-allocation.pdf [doi: 10.1108/09504120310455975]
54. Chang KR, Lou X, Karaletsos T, Crosbie C, Gardos S, Artz D, et al. bioRxiv - the preprint server for Biology. 2016. An Empirical Analysis of Topic Modeling for Mining Cancer Clinical Notes URL: https://www.biorxiv.org/content/10.1101/062307v1 [accessed 2020-01-07]
55. Cohen R, Aviram I, Elhadad M, Elhadad N. Redundancy-aware topic modeling for patient record notes. PLoS One 2014;9(2):e87555 [FREE Full text] [doi: 10.1371/journal.pone.0087555] [Medline: 24551060]
56. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed Representations of Words and Phrases and their Compositionality. In: Proceedings of the Advances in Neural Information Processing Systems 26. New York, NY: Curran Associates, Inc; 2013 Presented at: NIPS'13; December 5-10, 2013; Lake Tahoe, Nevada, USA p. 3111-3119 URL: https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf
57. Glicksberg BS, Miotto R, Johnson KW, Shameer K, Li L, Chen R, et al. Automated disease cohort selection using word embeddings from Electronic Health Records. Pac Symp Biocomput 2018;23:145-156 [FREE Full text] [Medline: 29218877]
58. Choi Y, Chiu CY, Sontag D. Learning low-dimensional representations of medical concepts. AMIA Jt Summits Transl Sci Proc 2016;2016:41-50 [FREE Full text] [Medline: 27570647]
59. Wang Y, Liu S, Afzal N, Rastegar-Mojarad M, Wang L, Shen F, et al. A comparison of word embeddings for the biomedical natural language processing. J Biomed Inform 2018 Nov;87:12-20 [FREE Full text] [doi: 10.1016/j.jbi.2018.09.008] [Medline: 30217670]
60. Kim Y. arXiv e-Print archive. 2014. Convolutional Neural Networks for Sentence Classification URL: http://arxiv.org/abs/1408.5882 [accessed 2020-01-07]
61. Baeza-Yates R, Ribeiro-Neto B. Modern Information Retrieval: The Concepts and Technology behind Search. Second Edition. Boston, MA: Addison-Wesley Professional; 2011.
62. Holzinger A, Biemann C, Pattichis CS, Kell DB. arXiv e-Print archive. 2017. What Do We Need to Build Explainable AI Systems for the Medical Domain? URL: http://arxiv.org/abs/1712.09923 [accessed 2020-01-07]
63. Lipton ZC. arXiv e-Print archive. 2016. The Mythos of Model Interpretability URL: http://arxiv.org/abs/1606.03490 [accessed 2020-01-07]
64. Marewski JN, Gigerenzer G. Heuristic decision making in medicine. Dialogues Clin Neurosci 2012 Mar;14(1):77-89 [FREE Full text] [Medline: 22577307]
65. Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with medical decisions: a systematic review. BMC Med Inform Decis Mak 2016 Nov 3;16(1):138 [FREE Full text] [doi: 10.1186/s12911-016-0377-1] [Medline: 27809908]
66. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is All you Need. In: Proceedings of the Advances in Neural Information Processing Systems 30. New York, NY: Curran Associates, Inc; 2017 Presented at: NIPS'17; December 4-9, 2017; Long Beach, CA, USA p. 5998-6008 URL: https://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf
67. Devlin J, Chang MW, Lee K, Toutanova K. arXiv e-Print archive. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding URL: http://arxiv.org/abs/1810.04805 [accessed 2020-01-07]
68. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov R, Le QV. arXiv e-Print archive. 2019. XLNet: Generalized Autoregressive Pretraining for Language Understanding URL: http://arxiv.org/abs/1906.08237 [accessed 2020-01-07]

Abbreviations

1D: 1 dimensional
AUC-PRC: area under the precision-recall curve
AUC-ROC: area under the receiver operating characteristic curve
BoN: bag of n-grams
ConvNet: convolutional neural network-based architecture
EHR: electronic health record
ICD-10: International Classification of Diseases, 10th revision

LASSO: least absolute shrinkage and selection operator
LBP: low back pain
LR: logistic regression
NLP: natural language processing
PCP: primary care provider
ReLu: rectified linear unit
RTW: return-to-work
TF-IDF: term frequency-inverse document frequency

Edited by C Lovis; submitted 01.11.19; peer-reviewed by V Osmani, M Boukhechba; accepted 15.12.19; published 27.02.20.

Please cite as:
Miotto R, Percha BL, Glicksberg BS, Lee HC, Cruz L, Dudley JT, Nabeel I
Identifying Acute Low Back Pain Episodes in Primary Care Practice From Clinical Notes: Observational Study
JMIR Med Inform 2020;8(2):e16878
URL: http://medinform.jmir.org/2020/2/e16878/
doi:10.2196/16878
PMID:32130159

Original Paper

A Communication Infrastructure for the Health and Social Care Internet of Things: Proof-of-Concept Study

Vincenzo Della Mea[1], MSc, PhD; Mihai Horia Popescu[1], MSc; Dario Gonano[2], MEng; Tomaž Petaros[3], MEng; Ivo Emili[3], MEng; Maria Grazia Fattori[2], MSc

[1]Department of Mathematics, Computer Science and Physics, University of Udine, Udine, Italy

[2]Cimtech Srl, Reana del Rojale, Italy

[3]MIPOT SpA, Cormons, Italy

Corresponding Author:
Vincenzo Della Mea, MSc, PhD
Department of Mathematics, Computer Science and Physics
University of Udine
Via delle Scienze 206
Udine, 33100
Italy
Phone: 39 0432 558461
Email: vincenzo.dellamea@uniud.it

Abstract

Background: Increasing life expectancy and reducing birth rates indicate that the world population is becoming older, with many challenges related to quality of life for old and fragile people, as well as their informal caregivers. In the last few years, novel information and communication technology techniques generally known as the Internet of Things (IoT) have been developed, and they are centered around the provision of computation and communication capabilities to objects. The IoT may provide older people with devices that enable their functional independence in daily life by either extending their own capacity or facilitating the efforts of their caregivers. LoRa is a proprietary wireless transmission protocol optimized for long-range, low-power, low–data-rate applications. LoRaWAN is an open stack built upon LoRa.

Objective: This paper describes an infrastructure designed and experimentally developed to support IoT deployment in a health care setup, and the management of patients with Alzheimer's disease and dementia has been chosen for a proof-of-concept study. The peculiarity of the proposed approach is that it is based on the LoRaWAN protocol stack, which exploits unlicensed frequencies and allows for the use of very low-power radio devices, making it a rational choice for IoT communication.

Methods: A complete LoRaWAN-based infrastructure was designed, with features partly decided in agreement with caregivers, including outdoor patient tracking to control wandering; fall recognition; and capability of collecting data for further clinical studies. Further features suggested by caregivers were night motion surveillance and indoor tracking for large residential structures. Implementation involved a prototype node with tracking and fall recognition capabilities, a middle layer based on an existing network server, and a Web application for overall management of patients and caregivers. Tests were performed to investigate indoor and outdoor capabilities in a real-world setting and study the applicability of LoRaWAN in health and social care scenarios.

Results: Three experiments were carried out. One aimed to test the technical functionality of the infrastructure, another assessed indoor features, and the last assessed outdoor features. The only critical issue was fall recognition, because a slip was not always easy to recognize.

Conclusions: The project allowed the identification of some advantages and restrictions of the LoRaWAN technology when applied to the health and social care sectors. Free installation allows the development of services that reach ranges comparable to those available with cellular telephony, but without running costs like telephony fees. However, there are technological limitations, which restrict the scenarios in which LoRaWAN is applicable, although there is room for many applications. We believe that setting up low-weight infrastructure and carefully determining whether applications can be concretely implemented within LoRaWAN limits might help in optimizing community care activities while not adding much burden and cost in information technology management.

(JMIR Med Inform 2020;8(2):e14583) doi:10.2196/14583

KEYWORDS

health services for the aged; remote sensing technology; sensors and actuators; embedded systems; Internet of Things; LoRaWAN

Introduction

The Scenario

Increasing life expectancy and reducing birth rates indicate that the world population is becoming older, with many challenges related to quality of life and well-being for old and fragile people, as well as their informal caregivers, particularly when functional independence decreases and there is a need to provide care on a daily basis. One of the conditions associated with most of the assistance burden is dementia.

In Italy, about 1.2 million people present with some form of dementia, and of these, about half are diagnosed with Alzheimer disease [1]. Patients may experience different levels of cognitive impairment. In the initial stages, they often are free to move around; however, they may not always be able to find their way. In fact, getting lost behavior is present in about 40% of patients with Alzheimer disease [2].

In previous years, Global Positioning System (GPS)–based technologies have been used to develop systems to support patients with Alzheimer disease and dementia in the stages of moderate impairment [3,4] by allowing caregivers to track them. Smartphones can be used for this approach; however, battery capacity and coverage have been reported as issues [3]. Systems have also been developed to work inside buildings [5]. Data collected from tracking systems appear useful to evaluate the evolution of the disease (eg, measure life space [6] and evaluate gait and balance [7]). Furthermore, recent systems allow tracking of people and, in principle, may provide guidance and support to moderately impaired patients [8].

In the last few years, novel information and communication technology techniques generally known as the Internet of Things (IoT) have been developed, and they are centered around the provision of computation and communication capabilities to objects, including those of common usage in daily life. By referring to the above-mentioned scenario, the IoT may provide older people with IoT objects that enable their functional independence in daily living by either extending their own capacity or facilitating the efforts of their caregivers while preserving and increasing, if possible, their functional independence in activities of daily living.

This aim could be achieved through an integrated infrastructure of smart objects with sensors, actuators, and intelligence that populate the homes of older community-dwelling individuals for their own use or caregiver use. These smart objects may include activity/presence sensors, vital sign readers, domotic actuators, and position trackers and may be based on different technologies like traditional sensors, as well as vision-based activity classifiers.

Internet of Things Requirements

The IoT is a paradigm used to describe “a variety of things or objects, such as radio-frequency identification tags, sensors, actuators, mobile phones, etc, which, through unique addressing schemes, are able to interact with each other and cooperate with their neighbors to reach common goals” [9]. This will be possible only when devices intercommunicate or at least communicate with some application able to receive and eventually send data.

For both data processing and communication, available resources might not always be sufficient, particularly regarding power and bandwidth. In fact, although some IoT objects can be connected to the electric grid as they are indoors and static, some others may be isolated and may require a battery. In this case, battery optimization and durability are strict requirements to guarantee long device backup. Consumption is dependent on processing power, sensors, and communication; thus, optimization needs to address these areas.

As the main aspect of this article is communication, we will focus on the communication infrastructure. In recent years, there has been rapid growth in proposals for low-power wide-area technologies. Although some very low consumption protocols already exist (Bluetooth Low Energy and ZigBee), their low consumption is achieved by substantially restricting the coverage of the transmitter. On the other hand, some IoT applications are deployed in wide areas, namely smart cities, fields, etc, where short range is not a viable option.

There are many technologies and standard proposals that address these needs, and the following three are prominent but different in their approaches:

- Sigfox: It is a network operator that manages a proprietary network based on unlicensed bands (eg, 868 MHz in Europe, 915 MHz in America, 433 MHz in Asia, and 915/923 MHz in Australia). It currently covers about 60 countries. There is no need for deployment and own infrastructure. However, if there is no coverage, it cannot be used. The business model is similar to that of cellular networks. There is a fee for using the network. Additionally, there are limits on message length and the daily number of messages sent. Moreover, the data rate is very limited (100 bits/s); however, this allows for a very long range (>40 km). As it uses free bands, quality of service is not guaranteed [10].
- LoRaWAN: It uses the same bands as Sigfox, but there is no central network operator; thus, there is no use fee [11]. Existing networks could be used or developers could create their own infrastructure. Although limits on data transfer are present, they are less strict than those in Sigfox, and the data rate varies depending on the spread factor (SF; 300 bits/s to 50 kilobits/s). The range is shorter than that of Sigfox (up to 20 km). Similar to Sigfox, quality of service is not guaranteed.
- NBIoT: It is the IoT technology of cellular network operators, exploiting the same licensed bandwidth and infrastructure as phones, with a similar business model [12]. It allows for guaranteed quality of service at a cost. The message length and data rate are substantially higher than

those of Sigfox and LoRaWAN. However, the node distance from the base station is less than 10 km.

Although all these technologies are made for the IoT, we selected LoRaWAN because of its ease of implementation, flexibility, and cost, as documented in a previous report [13], where a detailed comparison from many points of view can be found.

LoRa and LoRaWAN

LoRa is a proprietary wireless transmission protocol that is optimized for long-range, low-power, low–data-rate applications and is developed by Semtech (Camarillo, California) [14]. LoRa uses a proprietary spread spectrum modulation derived from chirp spread spectrum technology [15]. This allows LoRa to increase sensitivity by selecting the amount of spread used according to the radio parameter SF, in the range of 7 to 12. With an increase in sensitivity, the data rate decreases but longer distances can be reached. In addition, forward error correction is implemented, and this further improves robustness against noise. With an allocated bandwidth of 125 kHz, the data rate ranges from 250 to 5470 bits/s, corresponding to a minimum received signal strength indicator (RSSI) of −135.5 to −122.5 and a maximum payload size per packet of 59-230 bytes. This makes it clear that LoRaWAN is not made for transferring large amounts of data.

The frequency range in which LoRa operates has limits that often depend on national regulations, but these are typically aimed at limiting frequency occupation both in terms of transmission time and power. This is because being an unlicensed band, there is no restricted access, unlike in cellular telephony. For example, in Europe, the maximum transmission power is 14 dBi and the maximum duty cycle is between 1% and 10% depending on the specific frequency. National or international regulations specify other limits within unlicensed bands that will not be dealt with in this paper.

In a country where it is mandatory to respect the duty cycle limit, using a low data rate limits the number of packets that can be sent. Power consumption also depends on transmission time, so it is typically advised to select the highest data rate at which transmission succeeds. It should be noted that a LoRa receiver is able to discriminate parallel transmissions at different SFs, thus allowing better exploitation of the band.

LoRaWAN is an open stack built upon LoRa, which defines the communication protocol and system architecture for the network [11]. LoRaWAN takes into account different aspects including the following:

- Security: Packets are encrypted using AES128, with two ways of joining the network (over-the-air activation and activation by personalization), and there is frame counter management to enforce security.
- Optimization of airtime: In LoRaWAN, the network server may recognize when the node is transmitting at a nonoptimal data rate (according to RSSI) and send commands to the node to switch to a faster rate, with a function called adaptive data rate, or to decrease the power used for transmission.
- Application management: The network server associates a node to an application and forwards packets to it, eventually performing some conversion from raw bytes to JSON or another format.

These functions involve a cost to packet length. LoRaWAN adds a header of 13 bytes to be able to deal with the above aspects and sends extra downlink packets to implement the adaptive data rate. For example, an 8-byte payload needs to be added to the 13-byte LoRaWAN header, resulting in a total of 21 bytes.

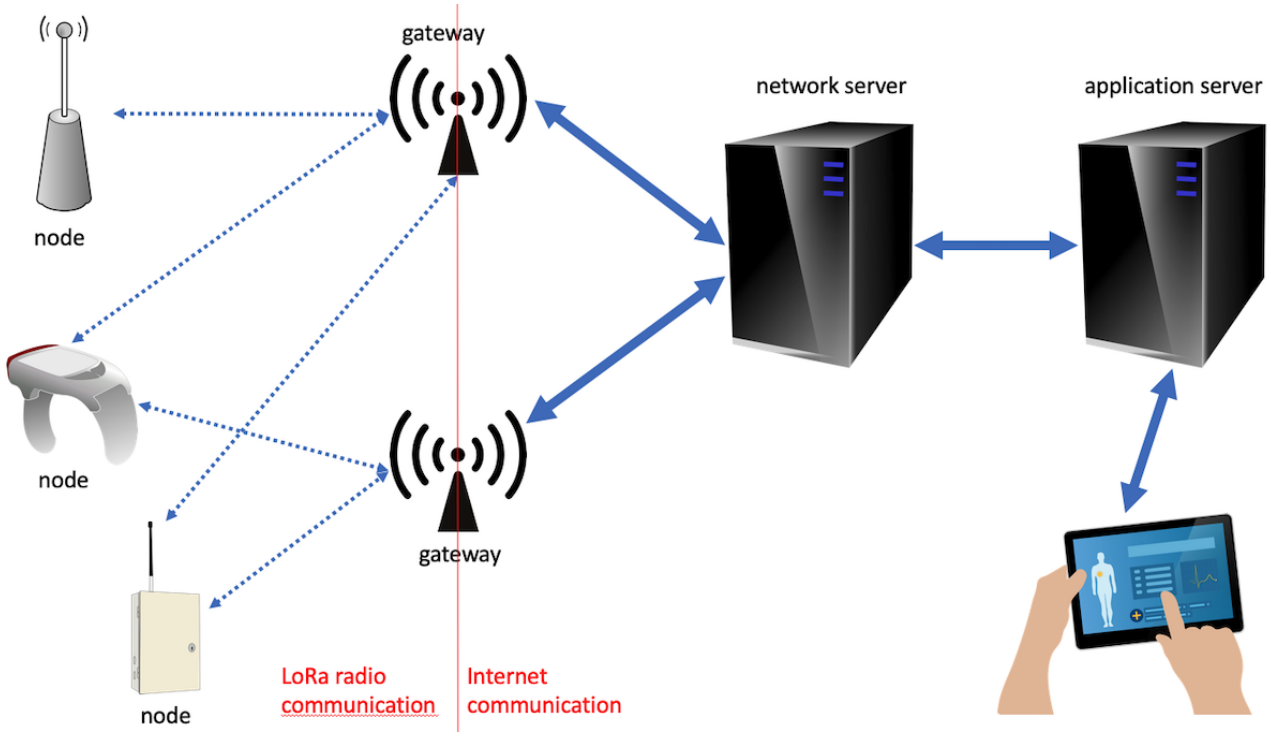
LoRaWAN defines data rates as specific configurations of SF and bandwidth on a regional basis and depending on the band. The bands are defined in terms of region and central frequency in MHz of the unlicensed spectrum. For example, EU868 refers to the European Region and frequencies ranging from 867.1 MHz to 869.525 MHz. In the European Region, the above-mentioned 21-byte packet will need from 28 ms (at SF7) to 1482 ms (at SF12) of airtime for transmission.

The typical architecture of LoRaWAN implementation is based on the following components:

- Node: It is a physical device, possibly with sensors, which can send and eventually receive packets to and from one or more gateways. In fact, nodes are not tied to a single gateway.
- One or more gateways: It principally receives LoRaWAN packets from nodes and forwards them through the Internet to a network server in a standardized format. It should be noted that nodes are not associated with a specific gateway, and their packets can be received by any gateway in the network.
- Network server: It receives packets from gateways, recognizes duplicate packets, requests nodes to change their data rate, decides to which application packets are directed, may perform some transformation of the payload, and finally sends packets to the selected application.
- One or more applications running on application servers that receive data from the network server.

Figure 1 shows the interaction among these components.

Figure 1. Interactions among LoRaWAN components.



As low power consumption is an important requirement, LoRaWAN nodes can be of three types, with functionality constraints that influence the power needed, as follows:

- Class A: Nodes mostly send packets and are able to receive only just after having sent a packet in two receive windows defined by the LoRaWAN standard (ie, 1 s and 2 s after delivery).
- Class B: Nodes may receive data in further slots of time defined and communicated by the network server through the gateway.
- Class C: Nodes are constantly listening for packets.

The most used nodes are Class A, because they represent the typical sensors sending data to an application, with no or limited need for receiving data.

The reported applications of LoRaWAN include agriculture [16], smart cities [17], environmental monitoring [18], and asset tracking and monitoring [19], and in general, it can be used for any form of remote monitoring.

Regarding the health domain, only few previous reports can be found, and most are from conference proceedings. Catherwood et al have described a portable diagnostic reader for urinary tract infection diagnosis that has been connected by means of LoRa [20]. Buyukakkaslar et al have investigated LoRaWAN as a possible electronic health technology but without addressing a specific scenario [21]. Peta¨ja¨ja¨rvi et al have framed their performance analysis in a scenario involving a person moving inside a confined and relatively small area both outdoors and indoors [22]. Mdhaffar et al. have performed a similar evaluation but in a larger geographical area [23]. The latter two papers provide useful insights for our experimentation.

Objectives

This paper describes a LoRaWAN-based infrastructure designed and experimentally developed to support IoT deployment in a health care setup, for which the management of patients with Alzheimer disease and dementia has been chosen as the experimentation field. The designed infrastructure covers all aspects, from physical devices to the application layer, partly exploiting existing technology and developing new modules. With the specific field of dementia management in mind, we developed a wearable device that is able to cover useful functions like localization and fall detection and a Web application for management, patient association, caregiver assignment, etc. The overall system is designed to be easily administered by both specialized and nonspecialized personnel, such as retirement home and hospital employees and patient relatives.

In the below sections, we will provide details on the technologies adopted in this project.

Methods

Project Aims

The above-mentioned infrastructure has been adopted in an industrial research project funded by the European Regional Development Program and Regional Operational Program framework through the Regional Government of Friuli-Venezia Giulia. The project entitled “Localization platform for people with cognitive disorders and dementia (PollicIoT)” had the following three main aims: (1) develop a complete system for outdoor patient localization from the hardware to the management platform; (2) ensure that the system can recognize and notify falls; and (3) ensure that the system has the potential for further clinical studies regarding the relationship between motion features and disease stage.

The project partners included an IoT system integrator (Cimtech, Reana del Rojale, Italy; coordinator), a hardware developer (MIPOT SpA, Cormons, Italy), the Medical Informatics & Telemedicine Lab at the University of Udine (responsible for platform design and data analysis), and a social care public company acting as a reference user group (ASP Moro, Codroipo, Italy).

During the project, the following further requirements were suggested by caregivers: (4) provide surveillance for patients subjected to night wandering and (5) provide indoor localization when patients are hosted in large structures.

Although each project aim is not novel by itself and has been implemented in some other system, in this case, the novelty lies in the communication infrastructure used to integrate all approaches.

The design process always involved all technical partners and needed the final users every time. The following sections provide details on the developed components.

The Node

The device given to patients was designed around a LoRaWAN chip previously developed by one of the project partners (MIPOT SpA). The node includes a GPS and an accelerometer with a barometric altimeter, and it was enriched with flash memory and a Bluetooth transceiver to fulfill the requirements (3) and (5), respectively.

Among the specific requirements for the node, small size and low weight were crucial, because it had to be worn by patients without disturbance. The final size is about 45×87×14 mm and weight is 55 g, and it includes a lithium polymer battery that could theoretically provide up to 1 week of use. Another specific requirement was to have the external surface as simple and anonymous as possible, avoiding buttons if possible. Thus, the node is black and smooth, with no light-emitting diodes and no switches. Its status can be recognized only through the platform, and it starts and remains on when charged. The prototype is charged through universal serial bus.

Figure 2 shows the node prototype. According to the LoRaWAN specifications for moving nodes, the node was set to send packets at a fixed SF of 12, which is the slowest setting, but it provides extended range. This value could be reconsidered depending on the range of the patient. For the experiment, the device was given to patients after agreement with their relatives, and it was worn by means of an armband specifically prepared by the retirement home personnel.

Figure 2. The PollicIoT wearable device.

To follow-up on a specific request of the users (requirement 4: night surveillance), towards the end of the project, we implemented a motion detector node, where the LoRaWAN transceiver is installed with an infrared sensor. This node can be placed in the rooms of patients who are known to wander at night, and it seamlessly integrates with the already available infrastructure.

The Internet of Things Infrastructure

Two gateways were installed for the project. One was placed at the coordinator site for the first functionality test and further roaming tests, and another was placed at the retirement home of the final users.

Regarding the network server, different platforms are already available on the market, and they almost always provide a free tier with some limitations and paid versions with more features or performance. The choice was The Things Network [24], which is a recent yet very active project based on an open source core that, in principle, could be directly used, although it does not have a graphical interface. Initial tests were conducted on the community edition, which is free but has a fair access policy that restricts the time-on-air per device and the number of downlink messages. The final implementation was flawlessly moved to the commercial cloud version.

The platform facilitates the integration of applications by providing a way to decode the binary payload sent by nodes to an application-specific JSON format, which is sent directly to a user-specified URL identifying the application server where the application is running.

The Application

The user interface of the PollicIoT system involves a Web platform. The core of the system receives JSON messages from

the IoT platform, and based on the messages, it localizes the associated patients on a map, sends alarms, etc.

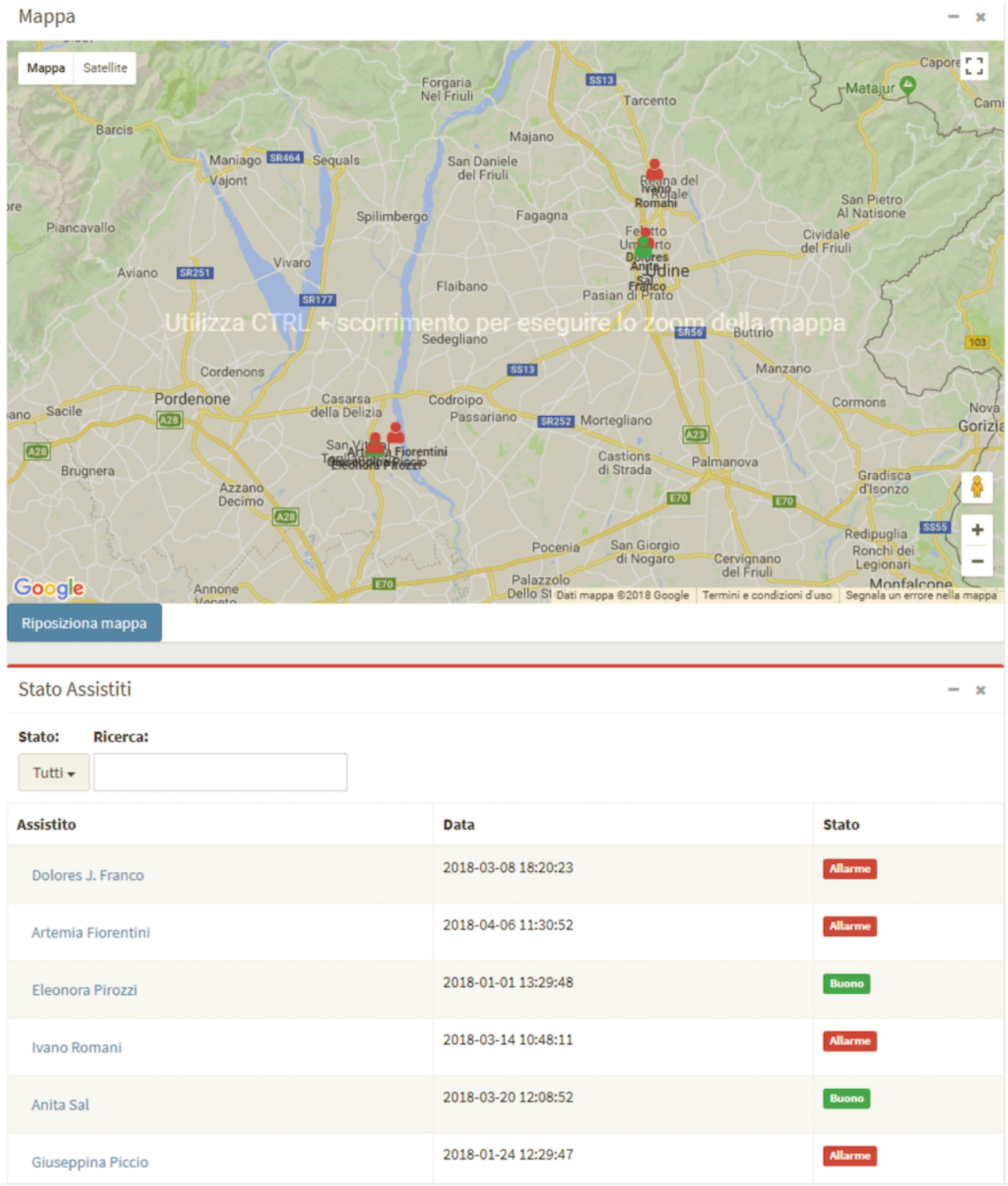
The following features have been implemented:

- Two-level geofencing: Alerts are sent whenever a patient exits one or more safety zones. These zones can be directly drawn on a map inside the platform and can be at the level of the structure hosting the patient (thus valid for every tracked patient) or specific for a single patient (eg, his/her relative's home and surroundings).
- Multichannel alerts: Alerts about geofencing, falls, battery exhaustion, etc, can be received on the Web platform, as well as through short message service messages and Telegram. The latter functionality has been implemented by means of a Telegram bot.
- Flexible structure management: The system may host one or more organizations, each of which may independently define one or more structures, corresponding to buildings, services, etc. Caregivers belong to a structure, and this allows direction of alerts to the most appropriate caregivers.

Figures 3 and 4 show screenshots of two sections of the Web application.

Figure 3. Screenshot of the geofence editor.

Figure 4. Screenshots of the surveillance map and patient status list.



Results

Technical Functionality

Three experiments were carried out. The first experiment was related to the evaluation of the technical functionality of the device and the infrastructure. For verification, two authors of this paper used the prototype device for more than a month, freely moving around, including travelling by car, although there is a report on the speed limit under which LoRaWAN is known to function at its best [25]. This allowed demonstration of the robustness of the involved systems and measurement of the maximum distance that could be reached from the gateways (up to 30.5 km in our experiment).

Thereafter, two experiments were aimed at evaluating the following two main application scenarios for patients still living at home: inside a community structure and in the town around the structure.

Residential Care Scenario

The former scenario involved a small number of patients who carried the wearable device inside a retirement home and nurses who accessed the Web platform for checking positions and alarms, with most movements made within the building. The experiment highlighted that when the device is far from a window, the GPS is not able to provide a position. This was foreseen, and the developed hardware was already designed to have Bluetooth-based localization for the indoors, although the necessary infrastructure was not developed because it was outside of the project.

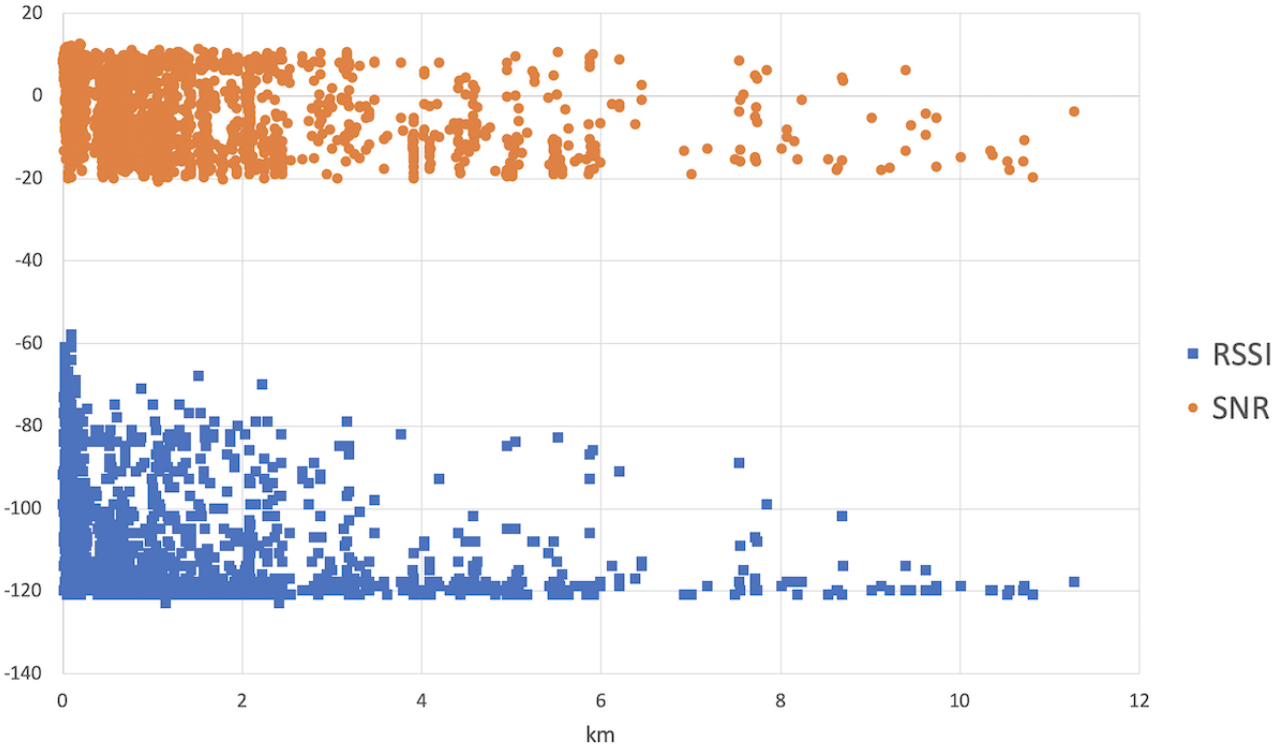
The fall recognition component was based on the free fall library made available by the company producing the accelerometer chip (STMicroelectronics, Geneva, Switzerland). Although the library can be customized by means of two thresholds (acceleration and altitude before and after the event), it does not appear to adequately capture the kinds of falls elderly people experience, including sliding from chairs, wheelchairs, and beds. In fact, two physiotherapists simulated typical falls while wearing the device, and the fall recognition module was not able to provide reliable alerts. Even the publicly available fall dataset [26] had data recorded during simulated falls among young people; thus, it might be difficult to use the data to develop a reliable elderly fall recognition algorithm. As the collection of true data about elderly falls is extremely difficult, the development of a new fall recognition module has been postponed.

Home Care Scenario

In the latter scenario, seven social workers carried around the wearable device to map the coverage of the gateway in their work area. This allowed the finding that nodes can communicate in a radius of up to 12 km around the gateway, which was considered sufficient to cover most of the patients in need. The radius obtained does not represent the maximum distance reachable by the nodes, because it is limited by the work area of the social workers, but the finding suggests the feasibility of using LoRaWAN in a typical setting. The map of points has not been included to maintain the privacy of the individuals involved.

Figure 5 shows the RSSI and signal-to-noise ratio (SNR), which are measures of signal intensity and quality, plotted according to the distance from the gateway and obtained from 3149 points collected during social worker trips. The plotted points were selected from a total of 3165 points, excluding those that were received by the other gateway in the network.

Figure 5. The received signal strength indicator (RSSI) and signal-to-noise ratio (SNR) of the packets received by the gateway.



The variabilities of the RSSI and SNR are high, because they depend on not only distance but also other geographical and environmental factors that cause path loss. The maximum distance recorded in this experiment and toward the main gateway was 11.27 km, although some packets were received by another gateway at 18.63 km.

Discussion

This project allowed the identification of some advantages and restrictions of the LoRaWAN technology when applied to the health and social care sectors.

Free installation allows the development of services that reach ranges comparable to those available with cellular telephony, but without the need for managing and paying for

communication. Although some of these services can be implemented using Wi-Fi inside a retirement home, they cannot be implemented outside, and LoRaWAN appears to be a sensible choice, particularly when dealing with elderly or disabled people who do not necessarily have an Internet connection at home. The following scenarios and applications could be envisaged:

- Social services could identify people in need of remote support, who do not have proper Internet infrastructure at home.
- Social services could identify the nodes needed. For example, a button for signaling events or requests [27]; a device that reminds about pill time; a wearable device that alerts in case of falls; a wearable device that collects data on vital signs like heart rate, steps, etc; and an activity monitor that signals the use of a television, coffee maker, etc.
- The needed nodes could be brought to the homes of patients, and they could be allowed to communicate without any extra infrastructure owned by the host.

Some of the functionalities described above, particularly collection of vital signs, could be implemented through Bluetooth-enabled devices that use LoRaWAN nodes as a bridge to the network, thus extending their original reach. However, to ensure full functionality of the described IoT solution, a survey of the transmission of nodes located indoors should be carried out, because the environment (wall size and composition, radiofrequency noise, etc) has a strong influence on radio transmission. This can be recognized by the findings presented in a report [22] for the indoors and another report [23] for the outdoors involving relatively large areas.

The indoor scenario had partial results, because indoor positioning via Bluetooth was not implemented, and this together with elderly fall recognition will be part of future work. LoRa, in principle, allows positioning by triangulation according to timestamps, with sufficient granularity for positioning precision in the order of meters.

Among the limits of the technology, two are crucial. First, with LoRaWAN, there is no assurance of real-time delivery of packets. In the case of conflicts during transmission, the packet might not be delivered in the first attempt, and it might or might not be resent depending on whether it is sent as “confirmed.” However, as confirmation is expensive in terms of network functioning, confirmed packets should be reduced to the minimum needed. Second, there is duty cycle constraint, which differs depending on country, but ranges between 0.1% and 10% maximum bandwidth occupation. Another limitation is the size of the packets, which is restricted like the bandwidth. All these limitations concur to set the following possible boundaries for LoRaWAN application in the health and social care areas:

- LoRaWAN cannot be used for emergency services, where true real-time communication is needed. This has been recognized in a previous report [20].
- LoRaWAN cannot be used to send large datasets, including images and full biosignal sets.
- LoRaWAN cannot be used when information needs to be sent at high frequencies, which in turn, depends on the distance. Nodes close to the gateway may send data at SF7, with a maximum frequency in the order of some seconds, whereas nodes far from the gateway have to send data at SF12, with a maximum frequency in the order of at least a couple of minutes.

These limits, excluding possibly the first one, can be overcome with smart nodes that embed intelligence to reduce transmission and packet dimension to the minimum needed for a specific application. For example, in our implementation of geofencing, the verification of patient position against the geofence is made at the server level; however, it could be performed inside the node, sending GPS positions only when outside the geofence, which would limit the number of packets sent. Another example is related to biosignals. They could be summarized by some indicator that is sent to the network application, and this could be eventually performed only when abnormal.

However, there are a number of use cases in which these limitations do not occur, including occasional alerts (falls, unforeseen movements, tracking during wandering, pressing communication buttons, etc [events that do not occur frequently]) and daily or generally timed collection of basic vital sign statistics like pressure, heartbeat, and activity, where transmission can be controlled and managed.

It should be noted that limitations are always present in best-effort wireless protocols, including Wi-Fi, because the band resource is shared. However, the potential of LoRaWAN for handling collisions and the capability of receiving signals at different SFs on the same frequency makes it relatively robust regarding issues, particularly with an increase in the number of available gateways. As prices are decreasing, extra gateways can easily be used to extend both the reach and reliability of communication. A fair number of cities and towns, particularly in Europe, are already covered by LoRaWAN public gateways as part of The Things Network [28], and there are other approaches available on other networks that do not disclose coverage maps. This infrastructure enables quick setup of feasibility studies, which can eventually be moved to private networks if needed.

Another crucial topic for health and social applications is security. Aras et al described a number of potential security attacks that were possible on v1.0 compliant LoRaWAN networks [29], and these are partly similar for all wireless technologies (ie, jamming techniques), partly linked to bad implementation practices (ie, no frame counter checks or use of activation by personalization), and partly related to physical tampering with the device. However, Butun et al carried out a similar analysis on version 1.1 of the LoRaWAN protocol and found that it addressed many of the security problems previously reported [30], although some new issues might have been introduced. Health and social care scenarios need to take into account these issues.

We believe that setting up low-weight infrastructure for the above-mentioned scenarios and carefully determining whether applications can be concretely implemented within LoRaWAN limits might help in optimizing community care activities while not adding much burden and cost in information technology management.

Acknowledgments

We thank the personnel at ASP Moro for their fruitful collaboration. The industrial research project “PollicIoT” was partially funded by the EU ERDF operational program through the government of the Friuli-Venezia Giulia region.

Conflicts of Interest

MGF is the owner and chief executive officer of Cimtech Srl, DG is the chief technical officer of Cimtech Srl, TP is the chief executive officer of MIPOT SpA, and IE is the research & development director at MIPOT SpA.

References

1. Prince M, Wimo A, Guerchet M, Ali G, Wu Y, Prina M. World Alzheimer Report 2015. London: Alzheimer's Disease International; 2015.
2. Pai M, Lee C. The Incidence and Recurrence of Getting Lost in Community-Dwelling People with Alzheimer's Disease: A Two and a Half-Year Follow-Up. PLoS One 2016 May 16;11(5):e0155480 [FREE Full text] [doi: 10.1371/journal.pone.0155480] [Medline: 27183297]
3. Faucounau V, Riguet M, Orvoen G, Lacombe A, Rialle V, Extra J, et al. Electronic tracking system and wandering in Alzheimer's disease: a case study. Ann Phys Rehabil Med 2009 Sep;52(7-8):579-587 [FREE Full text] [doi: 10.1016/j.rehab.2009.07.034] [Medline: 19744906]
4. Megges H, Freiesleben SD, Jankowski N, Haas B, Peters O. Technology for home dementia care: A prototype locating system put to the test. Alzheimers Dement (N Y) 2017 Sep;3(3):332-338 [FREE Full text] [doi: 10.1016/j.trci.2017.04.004] [Medline: 29067340]
5. Almudevar A, Leibovici A, Tentler A. Home monitoring using wearable radio frequency transmitters. Artif Intell Med 2008 Feb;42(2):109-120. [doi: 10.1016/j.artmed.2007.11.002] [Medline: 18215512]
6. Tung JY, Rose RV, Gammada E, Lam I, Roy EA, Black SE, et al. Measuring life space in older adults with mild-to-moderate Alzheimer's disease using mobile phone GPS. Gerontology 2014;60(2):154-162. [doi: 10.1159/000355669] [Medline: 24356464]
7. Hsu Y, Chung P, Wang W, Pai M, Wang C, Lin C, et al. Gait and Balance Analysis for Patients With Alzheimer's Disease Using an Inertial-Sensor-Based Wearable Instrument. IEEE J Biomed Health Inform 2014 Nov;18(6):1822-1830. [doi: 10.1109/jbhi.2014.2325413]
8. Pulido Herrera E. Location-based technologies for supporting elderly pedestrian in "getting lost" events. Disabil Rehabil Assist Technol 2017 May 04;12(4):315-323. [doi: 10.1080/17483107.2016.1181799] [Medline: 27377102]
9. Atzori L, Iera A, Morabito G. The Internet of Things: A survey. Computer Networks 2010 Oct;54(15):2787-2805. [doi: 10.1016/j.comnet.2010.05.010]
10. SIGGFOX. URL: https://www.sigfox.com/en [accessed 2019-05-03] [WebCite Cache ID 785d1xbIa]
11. LoRa Alliance Technical Committee. LoRaWAN 1.1 Specification. Beaverton, OR: LoRa Alliance, Inc; 2017.
12. Wang YE, Lin X, Adhikary A, Grovlen A, Sui Y, Blankenship Y, et al. A Primer on 3GPP Narrowband Internet of Things. IEEE Commun Mag 2017 Mar;55(3):117-123. [doi: 10.1109/MCOM.2017.1600510CM] [Medline: 27295638]
13. Mekki K, Bajic E, Chaxel F, Meyer F. A comparative study of LPWAN technologies for large-scale IoT deployment. ICT Express 2019 Mar;5(1):1-7. [doi: 10.1016/j.icte.2017.12.005]
14. Augustin A, Yi J, Clausen T, Townsley WM. A Study of LoRa: Long Range & Low Power Networks for the Internet of Things. Sensors (Basel) 2016 Sep 09;16(9) [FREE Full text] [doi: 10.3390/s16091466] [Medline: 27618064]
15. Simon MK, Omura J, Scholtz RA, Levitt BK. Spread Spectrum Communications Handbook. New York, NY: McGraw-Hill; 1994.
16. Davcev D, Mitreski K, Trajkovic S, Nikolovski V, Koteli N. IoT agriculture system based on LoRaWAN. 2018 Presented at: Inth IEEE International Workshop on Factory Communication Systems (WFCS), IEEE; 2018; Imperia, Italy p. 1-4. [doi: 10.1109/wfcs.2018.8402368]
17. Loriot M, Aljer A, Shahrour I. Analysis of the use of LoRaWan technology in a large-scale smart city demonstrator. : IEEE; 2017 Presented at: Sensors Networks Smart and Emerging Technologies (SENSET); 2017; Beirut, Lebanon p. 1-4. [doi: 10.1109/senset.2017.8125011]
18. Johnston SJ, Basford PJ, Bulot FM, Apetroaie-Cristea M, Easton NH, Davenport C, et al. City Scale Particulate Matter Monitoring Using LoRaWAN Based Air Quality IoT Devices. Sensors (Basel) 2019 Jan 08;19(1) [FREE Full text] [doi: 10.3390/s19010209] [Medline: 30626131]
19. Sanchez-Iborra R, G. Liaño I, Simoes C, Couñago E, Skarmeta A. Tracking and Monitoring System Based on LoRa Technology for Lightweight Boats. Electronics 2018 Dec 22;8(1):15. [doi: 10.3390/electronics8010015]
20. Catherwood PA, Steele D, Little M, Mccomb S, Mclaughlin J. A Community-Based IoT Personalized Wireless Healthcare Solution Trial. IEEE J Transl Eng Health Med 2018;6:2800313 [FREE Full text] [doi: 10.1109/JTEHM.2018.2822302] [Medline: 29888145]

21. Buyukakkaslar MT, Erturk MA, Aydin MA, Vollero L. LoRaWAN as an e-health communication technology. 2017 Presented at: IEEE 41st Annual Computer Software and Applications Conference (COMPSAC); 2017; Turin, Italy p. 310-313. [doi: 10.1109/compsac.2017.162]
22. Petäjäjärvi J, Mikhaylov K, Hämäläinen M, Iinatti J. Evaluation of LoRa LPWAN technology for remote health and wellbeing monitoring. 2016 Presented at: 10th International Symposium on Medical Information and Communication Technology (ISMICT); 2016; Worcester, MA p. 1-5. [doi: 10.1109/ismict.2016.7498898]
23. Mdhaffar A, Chaari T, Larbi K, Jmaiel M, Freisleben B. IoT-based health monitoring via LoRaWAN. 2017 Presented at: IEEE EUROCON 2017 - 17th International Conference on Smart Technologies; 2017; Ohrid, Macedonia p. 519-524. [doi: 10.1109/eurocon.2017.8011165]
24. The Things Network. URL: https://www.thethingsnetwork.org/ [accessed 2019-05-03] [WebCite Cache ID 785dJSbit]
25. Petäjäjärvi J, Mikhaylov K, Pettissalo M, Janhunen J, Iinatti J. Performance of a low-power wide-area network based on LoRa technology: Doppler robustness, scalability, and coverage. International Journal of Distributed Sensor Networks 2017 Mar 17;13(3):155014771769941-155014771769916. [doi: 10.1177/1550147717699412]
26. Casilari E, Santoyo-Ramón JA, Cano-García JM. Analysis of Public Datasets for Wearable Fall Detection Systems. Sensors (Basel) 2017 Jun 27;17(7) [FREE Full text] [doi: 10.3390/s17071513] [Medline: 28653991]
27. Chai PR, Zhang H, Baugh CW, Jambaulikar GD, McCabe JC, Gorman JM, et al. Internet of Things Buttons for Real-Time Notifications in Hospital Operations: Proposal for Hospital Implementation. J Med Internet Res 2018 Aug 10;20(8):e251 [FREE Full text] [doi: 10.2196/jmir.9454] [Medline: 30097420]
28. The Things Network Map. URL: https://www.thethingsnetwork.org/map [accessed 2019-05-03] [WebCite Cache ID 785dZFd9g]
29. Aras E, Ramachandran GS, Lawrence P, Hughes D. Exploring the Security Vulnerabilities of LoRa. 2017 Presented at: 3rd IEEE International Conference on Cybernetics (CYBCONF); 2017; Exeter, UK p. 1-6. [doi: 10.1109/cybconf.2017.7985777]
30. Butun I, Pereira N, Gidlund M. Analysis of LoRaWAN v1.1 Security. 2018 Presented at: 4th ACM MobiHoc Workshop on Experiences with the Design and Implementation of Smart Objects, Los Angeles; 2018; Los Angeles, CA p. 1-5. [doi: 10.1145/3213299.3213304]

Abbreviations

GPS: Global Positioning System
IoT: Internet of Things
RSSI: received signal strength indicator
SF: spread factor
SNR: signal-to-noise ratio

Edited by G Eysenbach; submitted 08.05.19; peer-reviewed by N Miyoshi, D Mendes, B Chaudhry; comments to author 03.10.19; revised version received 22.11.19; accepted 16.12.19; published 25.02.20.

Please cite as:
Della Mea V, Popescu MH, Gonano D, Petaros T, Emili I, Fattori MG
A Communication Infrastructure for the Health and Social Care Internet of Things: Proof-of-Concept Study
JMIR Med Inform 2020;8(2):e14583
URL: http://medinform.jmir.org/2020/2/e14583/
doi:10.2196/14583
PMID:32130158

Original Paper

Analysis of Massive Online Medical Consultation Service Data to Understand Physicians' Economic Return: Observational Data Mining Study

Jinglu Jiang[1], PhD; Ann-Frances Cameron[2], PhD; Ming Yang[3], PhD

[1]Binghamton University, Binghamton, NY, United States

[2]HEC Montreal, Montreal, QC, Canada

[3]Central University of Finance and Economics, Beijing, China

Corresponding Author:
Jinglu Jiang, PhD
Binghamton University
4400 Vestal Pkwy E
Binghamton, NY, 13902
United States
Phone: 1 6077773016
Email: jingluj@binghamton.edu

Abstract

Background: Online health care consultation has become increasingly popular and is considered a potential solution to health care resource shortages and inefficient resource distribution. However, many online medical consultation platforms are struggling to attract and retain patients who are willing to pay, and health care providers on the platform have the additional challenge of standing out in a crowd of physicians who can provide comparable services.

Objective: This study used machine learning (ML) approaches to mine massive service data to (1) identify the important features that are associated with patient payment, as opposed to free trial–only appointments; (2) explore the relative importance of these features; and (3) understand how these features interact, linearly or nonlinearly, in relation to payment.

Methods: The dataset is from the largest China-based online medical consultation platform, which covers 1,582,564 consultation records between patient-physician pairs from 2009 to 2018. ML techniques (ie, hyperparameter tuning, model training, and validation) were applied with four classifiers—logistic regression, decision tree (DT), random forest, and gradient boost—to identify the most important features and their relative importance for predicting paid vs free-only appointments.

Results: After applying the ML feature selection procedures, we identified 11 key features on the platform, which are potentially useful to predict payment. For the binary ML classification task (paid vs free services), the 11 features as a whole system achieved very good prediction performance across all four classifiers. DT analysis further identified five distinct subgroups of patients delineated by five top-ranked features: previous offline connection, total dialog, physician response rate, patient privacy concern, and social return. These subgroups interact with the physician differently, resulting in different payment outcomes.

Conclusions: The results show that, compared with features related to physician reputation, service-related features, such as service delivery quality (eg, consultation dialog intensity and physician response rate), patient source (eg, online vs offline returning patients), and patient involvement (eg, provide social returns and reveal previous treatment), appear to contribute more to the patient's payment decision. Promoting multiple timely responses in patient-provider interactions is essential to encourage payment.

(JMIR Med Inform 2020;8(2):e16765) doi:10.2196/16765

KEYWORDS

Web-based health services; remote consultation; machine learning; data mining; decision tree; patient involvement

Introduction

Background

Online health care solutions are increasingly popular [1-3], with reports that they are preferred by more than 70% of patients [4]. This study focuses on *multisided online medical consultation platforms* where various health care providers from different hospitals and medical institutes provide remote medical consultation services to patients. This type of digital health care service is experiencing significant growth and research attention [5]. These platforms offer many benefits, such as reduced medical costs, improved medical service efficiency, more efficient health care resource distribution, and fewer health care resource shortages in remote areas [2,6-9].

Despite the popularity and potential benefits, some online medical consultation platforms are struggling to attract and retain patients who are willing to pay for these services, for example, patient dissatisfaction after an initial failed experience, fear that diagnoses are made with limited consideration of patients' medical history, and concerns about privacy may impede patients' use of online consultation [10,11]. In addition, online medical consultation usually follows the Pareto principle in that 80% of the services are provided by 20% of the physicians on the platform [1], suggesting that many health care service providers on the platform have the challenge of attracting patients and standing out in the crowd of physicians who can provide comparable services [6,12]. To entice patients to their platform and promote payment, many platforms employ a multitiered pricing strategy that allows the coexistence of free (ie, the free trials) and paid versions (ie, the premium) of services [13]. As a consequence, patients may be more willing to pay for the service, and physicians may be able to access a broader range of patients.

Several features associated with patient payment in online medication consultation platforms have been frequently examined by previous research. Physician reputation—both online and offline—is the most frequently examined physician characteristic [14]. As medical consultation is highly professional, physicians need to be credible or trustworthy to attract and retain paying patients. A physician's affiliation, seniority, and location are usually used as proxies for reputation [8,9,15,16]. Patient evaluation, which is the feedback left on the platform by previous patients about the physician, is also frequently examined [2,6,17]. It is often displayed in the form of ratings, stars, reviews, and virtual gifts. This feedback is visible to other patients on the platform and may serve as signals of service quality, which impact patients' willingness to pay. Although less frequently examined, patient-physician interaction may be an important feature as well. The frequency and depth of interaction on the platform (eg, the amount of service or the frequency of service) show the ability and willingness of a physician to provide high-quality service, which may influence patient payment [3,15,18,19].

Gaps and Objectives

This existing research is useful; however, these service and physician-related features are often examined in isolation and often using a linear regression approach. Thus, the understanding of how various features interact to generate impacts is currently lacking—although some features might be important enough to generate impacts on their own, others may only have impacts when combined with other features. To extend existing research, new approaches are needed, which take advantage of the massive data on these platforms and help uncover the complex dynamics between these various features and their interactions and payment. Thus, the objectives of this study were to determine (1) the important features of online medical consultation services that are associated with patient payment, as opposed to free trial–only appointments; (2) the relative importance of these features; and (3) how these features interact, linearly or nonlinearly, in relation with payment. We focus on mining feature importance because knowing the features (and their interactions), which influence payment, will help platforms and physicians identify high-value online medical consultations. Although many features may impact payment, we are particularly interested in those, which are publicly visible on the platform, such as characteristics of physicians and their interaction with patients and patient feedback, rather than nonvisible features, such as patients' economic status and their general attitude toward technology. This is because publicly visible features contain information and signals that, through observational learning and social influence [20-23], may influence patient payment.

To this end, we examine a massive dataset from the largest China-based online medical consultation platform (1.5 million patient-physician consultation records) spanning 10 years. Predictive models are developed by employing classic machine learning (ML) procedures (ie, feature selection, hyperparameter tuning, model training, and validation) with logistic regression (LR), simple decision tree (DT), random forest (RF), and gradient boost (GB) classifiers. The importance ranking of these features is identified through regression coefficients, level of DT splits, and feature importance scores provided by RF and GB algorithms.

Methods

Empirical Setting and Dataset

Our empirical setting is a multisided online medical consultation platform based in China. It is one of the largest medical platforms, and more than half a million physicians from over 9400 hospitals have set up their profiles and provided consultation services on the platform. The platform follows a service model that allows the coexistence of free and paid consultation services (see Multimedia Appendix 1 for more details).

Our dataset includes 10 years of consultation records (approximately 2.3 million records from January 2009 to August 2018) between patient-physician pairs from three departments that have received the most visits (ie, pediatrics, gynecology, and dermatology, according to the platform report) across six geographic areas—three of the areas are those with the richest health care resources (Beijing municipality, Guangdong province, and Zhejiang province) and three are remote areas with the fewest health care resources (Shanxi province, Tibet province, and Qinghai province). Each record is a consultation

history that includes picture- and text-based dialogs and service purchase records between patient *i* and physician *j* (see Multimedia Appendix 1).

Machine Learning Task and Initial Feature Selection

Our focal outcomes are whether a consultation record includes payment and the relative importance of the features on the platform that can predict payment. Although a consultation record may include multiple times of payments, we do not consider payment intensity or types. Accordingly, the objective of our ML task is to solve a binary classification problem—classifying consultation records into free services only (labeled as *free*) or those including some type of financial payment (labeled as *paid*). The consultation with a *paid* label is our positive class in ML prediction.

The initial 18 features were identified by drawing on variables that have been examined in previous studies (see Table 1 for definition and coding of features) and were consistently visible on the platform. Features that are visible to platform users (eg, visitors, patients, and physicians) may influence payment, as they potentially allow patient learning and valuation to occur before the actual consumption of the consultation service. Although the importance of online physician reputation has been demonstrated in previous studies [19], physicians' online rating was not included in this study. Owing to the changes in platform design, online reputation scores (eg, stars, ratings, and reviews) are not consistent over time. In addition, we observed that most physicians have very good ratings with little variation (mean 3.80, SD 0.34), which would have made this feature less useful as a predictor. This ceiling effect has been reported in the previous study using the same context [1,9]. However, features such as social returns and service intensity were included and can reflect physicians' online reputation to some extent [9].

Table 1. Key predictive features and coding description.

Feature	Description	Reference
Physician reputation related		
Hospital ranking[a]	• [Ranking 1] Equals 1 if primary care hospital, 0 otherwise. • [Ranking 2] Equals 1 if secondary care hospital, 0 otherwise.	[2,3,9,14,19]
Physician seniority	• [Title 1] Equals 1 if chief physician, 0 otherwise. • [Title 2] Equals 1 if associate chief physician, 0 otherwise.	[2,3,9,14,19]
Hospital location	• [Loc] Equals 1 if health care resource–rich areas, 0 otherwise.	[9,15]
Physician tenure	• [Tenure] The number of months the physician has been registered on the platform.	[15,19]
Service intensity	• [Intensity] The average number of patients served per month during the physician's tenure (=total patients served/tenure).	[7]
Patient related		
Previous formal examination	A function provided by the platform allowing patients to reveal their medical status: • Status 1: no formal health care examination before the consultation. • Status 2: a formal health care examination before the consultation. • Status 3: private (ie, detailed consultation information is not directly visible by other patients). (coded into dummies) • [PriorExam] Equals 1 if none, 0 otherwise. • [Private] Equals 1 if set as private, 0 otherwise.	N/A[b]
Offline connection	• [Offline] A check-in function provided by the platform to indicate patients' offline connection with the physicians. Equals 1 if the patient used the check-in function, 0 otherwise.	[16]
Service delivery related		
Service duration	• [Duration] Number of days between the initial post and last post of patient i's interaction with physician j.	N/A
Total dialog	• [TotalD] Total number of posts within patient i's interaction with physician j.	[18]
Physician posts	• [PhysicianP] Number of posts initiated by physician j within patient i's interaction with physician j.	[3]
Response rate	• [Response] The rate of response of a physician (=PhysicianP/TotalP).	N/A
Answer frequency	• [Answer_frq] The average number of answers (including notifications and reminders) by the physician per day during patient i's interaction with physician j (=PhysicianP/Duration).	N/A
Social return	• [Social] A function provided by the platform to allow patients to send virtual gifts to the physician. Equals 1 if patient i gave any virtual gift to physician j at any time during patient i's interaction with physician j.	[2,3,12,15,18,19]
Patient involvement related		
Patient posts	• [PatientP] Number of posts initiated by patient i within that patient's interaction with physician j.	[3]
Question frequency	• [Question_frq] The average number of posts by the patient per day during patient i's interaction with physician j (=PatientP/Duration).	N/A

[a]Hospital ranking in China is a three-tier system (primary, secondary, and tertiary institutions) based on the hospital's ability to provide medical care, education, and research; thus, physicians who have been able to secure a position at a primary care hospital are generally considered to be of higher reputation [24].

[b]Not applicable.

Data Cleaning and Analysis Pipeline

First, data were prepared by removing consultation records that did not fit the scope of the study (eg, consultation occurred before 2009 and after 2018 and samples with unqualified tags). We also excluded records with over 50% of missing values (N=84,582) and outliers using the 95% quantile as the threshold (N=674,767; see Multimedia Appendix 2 for a detailed description of data cleaning procedure).

In the second step, four data-driven feature selection techniques were applied to identify the right features to use in the ML classification (low variance filtering, high correlation filtering, backward feature selection, and forward feature selection) [25,26]. The objective of this procedure is to find the features that are highly correlated with the outcome but ideally uncorrelated with each other [27] so that the resulting features can build a relatively parsimonious model (see Multimedia Appendix 3 for a detailed description of feature selection procedure).

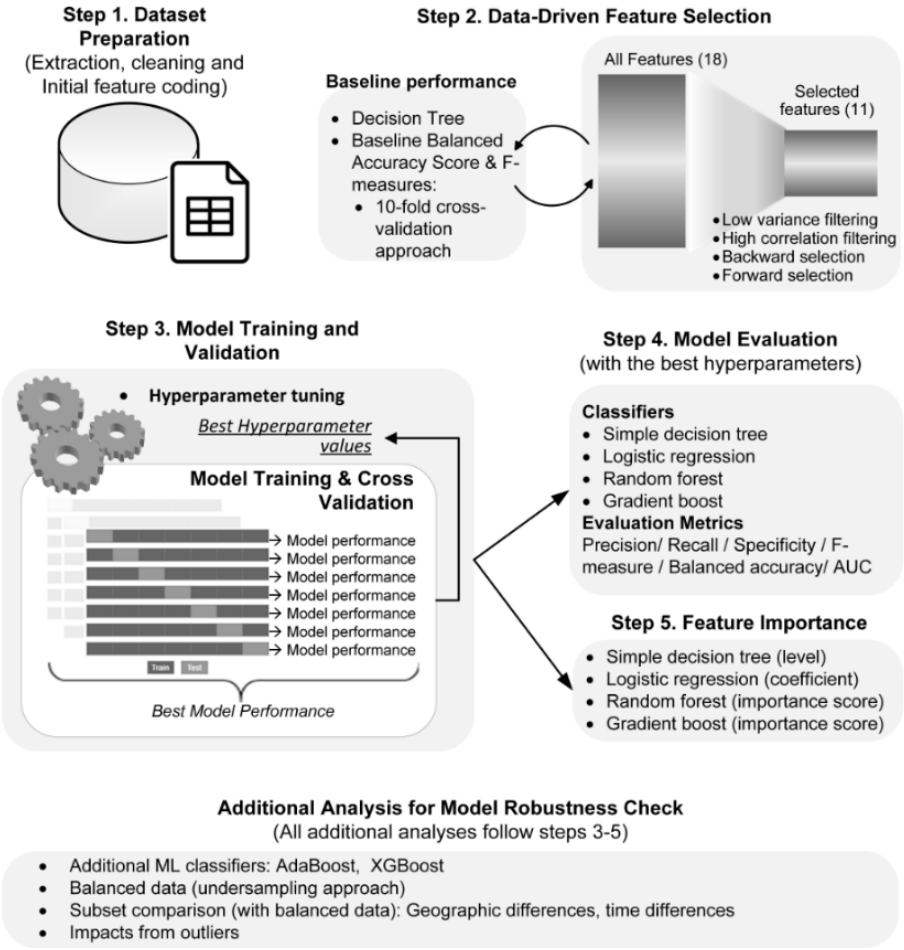
In step 3, the ML model was constructed through three nested procedures: hyperparameter optimization, model training, and validation (see Figure 1). Four common ML classifiers were purposefully chosen—LR, DT, RF, and GB—because they are mainstream ML techniques for classification problems [13] accessible by general data consumers through data analysis tools and platforms (eg, Python, R, SAS, and RapidMiner). LR was used in previous studies with small datasets [2], and the latter three are tree-based approaches with different resampling strategies and cost function optimization techniques (ie, boosting vs bagging and gradient descent algorithm). Depending on the ML classifier, different sets of hyperparameters need to be configured to ensure that the algorithm reaches its best classification performance (see Multimedia Appendix 2 for a detailed explanation of optimization and analysis procedures). We conducted our analysis on the KoNstanz Information MinEr platform.

The performances of the resulting ML models were compared in step 4. We used six evaluation metrics, which are commonly accepted in ML classification and can reflect different aspects of ML model performance (eg, correctly assign the paid services with a paid label vs the probability that an ML classifier will successfully classify a case in the right class) [28,29] (see Multimedia Appendix 2 for detailed explanation of our evaluation metrics).

We investigated research objectives 2 and 3 through step 5, which examines feature importance. The four classifiers that we used provide different feature importance indicators—the regression coefficients in LR, level of splits for DT, and feature importance indices for both GB and RF (see Multimedia Appendix 2, and the study by Friedman [30]).

For steps 3 to 5, there are some particularities of our data that may bias our results (eg, imbalanced data). Thus, we perform several additional tests to examine the robustness of the model. The results of these additional analyses indicate that our model is robust to sample distribution (eg, imbalances, classes, and outliers) and potential systematic differences (eg, geographic location and market changes), as indicated by only minor changes in the model performance measures (see Multimedia Appendix 4 for the results of these additional analyses).

Figure 1. Analysis pipeline. AUC: area under the receiver operating characteristic curve; ML: machine learning.



Results

Feature Selection Results and Descriptive Statistics

After data cleaning, 1,582,564 qualified records remained for further analysis. Among these records, 1,089,662 (68.85%) were free trial–only, whereas 492,902 (31.15%) involved at least one premium payment. After performing four feature selection techniques (step 2, see Multimedia Appendix 3), we retained the ones that are selected by forward, backward, low variance filtering, and high correlation filtering approaches. In response to our first research objective regarding which features of online medical consultation services are associated with patient payment, our feature selection analysis suggested 11 key features (see Table 2)—the seven eliminated features were thus considered as less useful because of either high correlation with the included features (ie, redundant features) or low variance explained (ie, low explanatory power).

Table 2. Summary statistics of features.

Service feature	All, mean (SD)	Free-only[a], mean (SD)	Paid[a], mean (SD)	All (minimum, maximum)	Free-only (minimum, maximum)	Paid (minimum, maximum)
Physician reputation related						
Hospital ranking 2	0.02 (0.15)	0.02 (0.16)	0.01 (0.12)	0, 1	0, 1	0, 1
Physician title 1	0.46 (0.5)	0.43 (0.5)	0.53 (0.5)	0, 1	0, 1	0, 1
Patient related						
PriorExam	0.19 (0.39)	0.06 (0.24)	0.47 (0.5)	0, 1	0, 1	0, 1
Private	0.08 (0.27)	0.07 (0.25)	0.1 (0.3)	0, 1	0, 1	0, 1
Offline connection	0.71 (0.45)	0.87 (0.33)	0.36 (0.48)	0, 1	0, 1	0, 1
Service delivery related						
Total dialog	7.44 (6.38)	6.27 (5.03)	10.04 (8.06)	1, 35	1, 31	1, 35
Response rate	0.19 (0.16)	0.18 (0.16)	0.2 (0.17)	0, 0.875	0, 0.75	0, 0.875
Answer frequency	0.22 (0.33)	0.24 (0.35)	0.18 (0.29)	0, 1.25	0, 1	0, 1.25
Social return	0.18 (0.38)	0.18 (0.38)	0.18 (0.29)	0, 1	0, 1	0, 1
Patient involvement related						
Patient posts	5.79 (5.10)	5.05 (4.38)	7.43 (6.10)	1, 28	1, 28	1, 28
Question frequency	1.11 (1.2)	1.2 (1.24)	0.92 (1.07)	0, 5.5	0, 5.5	0, 5.5

[a]Mean differences between free and paid services are all significant (P<.001), except for social return (P=.025).

Machine Learning Model Performance and Feature Importance Ranking

Next, the overall model performance was examined (step 4; see Table 3). As we have an imbalanced dataset (ie, the ratio between paid and free-only services is around 1:2), area under the receiver operating characteristic curve (AUC), F measure, and balanced accuracy are less biased and more informative than other measures. GB exhibited the best overall performance (balanced accuracy=0.973, F measure=0.97, and AUC=1). However, all classifiers performed well, indicating that our predictive model with 11 selected features exhibits significant classification performance. Explanation of each measure is presented in Multimedia Appendix 2.

In investigating our research objective on the relative importance of the 11 features, the four ML classifiers yielded relatively consistent results in the top-ranked and low-ranked features, whereas the ones in the middle were less consistent (Table 4). Offline connection, response rate, social return, total dialog, diagnoses from a prior examination, and private status consistently ranked high, whereas physician title, question frequency, and the second-tier hospital ranking were consistently ranked low.

Table 3. Machine learning model performance evaluation.

Model performance measurement	Logistic regression	Decision tree	Gradient boost	Random forest
Recall	0.851	0.949	0.952	0.908
Precision	0.896	0.989	0.988	0.984
Specificity	0.956	0.995	0.995	0.993
F measure	0.873	0.969	0.970	0.944
Balanced accuracy	0.903	0.972	0.973	0.951
Area under the receiver operating characteristic curve	1.000	0.988	1.000	0.988

Table 4. Key features listed in descending order of importance.

Service feature	Logistic regression (coefficient[a])	Decision tree (level of splits)	Gradient boost (importance, %)	Random forest (importance, %)
1	Response rate (−13.89)	Offline connection (1)	Offline connection (30)	Offline connection (24)
2	Offline connection (−4.99)	Social return (2)	Total dialog (30)	PriorExam (20)
3	Social return (−3.11)	Total dialog (2)	Response rate (25)	Total dialog (18)
4	Patient posts (−2.63)	Private (3)	Social return (8)	Response rate (17)
5	Total dialog (2.47)	Response rate (3)	Private (6)	Patient post (9)
6	PriorExam (1.70)	PriorExam (4)	Patient posts (1)	Social return (7)
7	Private (−0.99)	Answer_frq (4)	PriorExam (0)	Private (2)
8	Ranking 2 (−0.305)	Patient posts (6)	Answer_frq (0)	Answer_frq (2)
9	Answer_frq (−0.14)	Question_frq (6)	Question_frq (0)	Question_frq (1)
10	Question_frq (−0.13)	Ranking 2 (8)	Title1 (0)	Title1 (0)
11	Title1 (−0.089)	Title 1 (9)	Ranking 2 (0)	Ranking 2 (0)

[a]For logistic regression, a regularization procedure (see Multimedia Appendix 2) is applied, so large weight coefficients are penalized for avoiding overfitting. All coefficients are significant (P<.001).

Interpreting Key Patient Subcategories Based on Feature Configurations

To address the third research objective, we examined how these features interact in relation to patient payments. A tree structure was used because it explicitly displays the feature hierarchies and classification outcomes at each tree split. Five key feature configurations emerged, which describe five subgroups of patients who interact with physicians differently, yielding different payment outcomes. By applying the learned tree structure on the full dataset, these five subgroups covered 85.2% of the total population, using a combination of only four key features (ie, offline, total dialog, response rate, and social return). Note that the DT algorithm has the capability to fully classify the whole population (in our case, at 10 layers), but the configurations become complex and practically less useful. Thus, we used the subgroups up to the third layer (see Figure 2 and Table 5).

Figure 2. Decision tree for identifying patient subgroups with the full dataset.

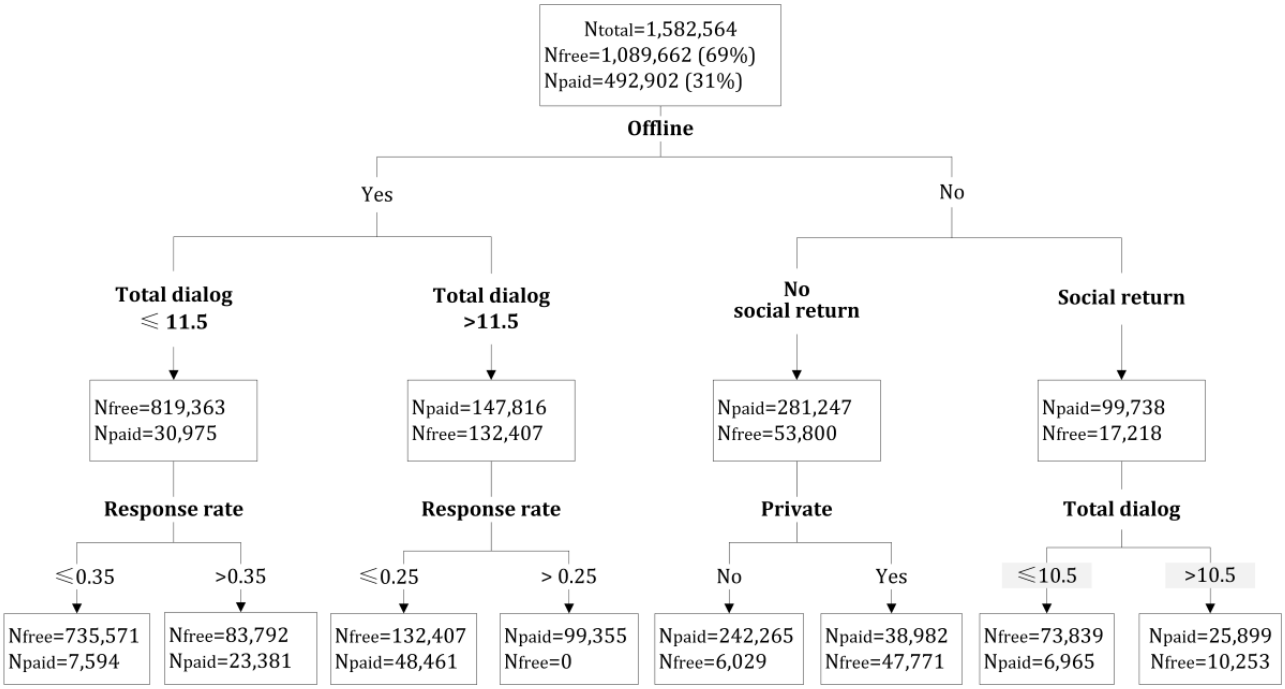


Table 5. Decision tree–based configuration of feature contributions.

Top feature configurations[a]	Number of cases in the node, n	Dominant outcome	Percentage of dominant cases, n (%)
Subgroup 1: These configurations suggest a simple type of follow-up service resulting from previous offline diagnoses			
Offline AND low total dialog (≤11.5)	850,338	Free	819,363 (96.36)
Offline AND low total dialog (≤11.5) AND low response rate (≤0.35)	743,165	Free	735,571 (98.98)
Subgroup 2: This configuration suggests a complex service extension from the previous offline diagnoses, which requires intensive patient-provider interaction.			
Offline AND high total dialog (≤11.5) AND high response rate (>0.25)	99,355	Paid	99,355 (100)
Subgroup 3: These configurations suggest the patient has no offline connection with the physician but is paying a premium for the online consultation rather than using a social return to show gratitude.			
Nonoffline AND no social return	335,047	Paid	281,247 (83.94)
Nonoffline AND no social return AND nonprivate	248,294	Paid	242,265 (97.57)
Subgroup 4: This configuration suggests the patient has no offline connection with the physician, has a less intensive online consultation experience, and offers a social return as compensation instead of payment.			
Nonoffline AND social return AND low total dialog (≤10.5)	80,804	Free	73,839 (91.38)
Subgroup 5: This configuration suggests the patient has no offline connection with the physician and engages in an intensive online interaction, providing both payment and a social return as compensation.			
Nonoffline AND social return AND high total dialog (>10.5)	36,152	Paid	25,899 (71.64)

[a]For each tree split, if no dominant outcome emerges (ie, free cases <80% or paid cases <70% at the focal split), we do not consider it as an important subgroup because additional service features are required to better classify these cases.

We can observe that patients who have previous offline consultations with the physician are less likely to pay. It is possible that these patients tend to take free opportunities to clarify simple unsolved issues after their offline visits, as indicated by increasing the proportion of free services in the presence of low total dialog and low response rates from the physicians (subgroup 1). However, if complex issues emerge, these patients may still prefer to return to the offline health care channel rather than pay for the premium online service.

A second type of returning patients (subgroup 2) may have complex issues and decide to stay online and pay. This represents a complex service extension: these returning patients may have complex issues that require highly interactive patient-physician communication. Thus, these returning patients frequently communicate with the physicians (probably because of the complexity of the issue) and receive frequent responses, which, in turn, are associated with a high probability of payment.

For online patients who have no prior connection with the physician, those who do not provide social returns (eg, thank you letters and virtual gifts) seem more likely to pay (subgroup 3). There may be a psychological compensation effect [31] where giving virtual gifts substitutes for the actual payment and balances the sense of *guilt* after receiving free services. However, in cases where the service between patients and physicians with no offline connection is highly interactive (ie, large amount of dialog), patients provide both virtual gifts and premium payment to show their appreciation (subgroup 4 vs subgroup 5).

The high-level presence of *private* in one of the tree branches deserves more attention. *Privacy* represents a function provided by the platform, which allows patients to set their dialogs as private, so they cannot be viewed by other people. From previous studies, we know that one of the major reasons that patients do not use online health care services is privacy concerns [32,33]. Patients who use this function may have a higher privacy concern than those who do not use it. As online medical consultation requires patients to reveal sensitive health-related information, patients who allow this information to be publicly displayed probably have lower privacy concerns and may be more likely to be more engaged in the online consultation and subsequent diagnosis. Owing to this heightened engagement, they may be more likely to pay after the initial free interactions (subgroup 3).

In summary, the source of patients (offline returning or online directly) seems to be a key differentiator for payment, which may be because of the different motivations and service requirements inherent in these two types of patients. Patient-physician interaction representing service delivery quality is another key differentiator (eg, total dialogs, response rate, and patient posts), which also indicates the importance of patient involvement and physician's timely response during the consultation. Privacy setting and social return, two features pertaining to the platform functionality, play important roles as well.

Discussion

Principal Findings

In this study, we focused on online medical consultation, a type of emerging digital health care service that has received much attention in recent years. Our objective was to understand the features of online medical consultation services that contribute to payment so that the platform can identify high-value services and take actions to better manage service providers and their offerings. As an initial study using ML approaches to identify key features and to make predictions, we did not aim to incrementally improve prediction accuracy by engineering the features or developing new algorithms. Rather, our goal was to develop a predictive model that has both sufficient explanatory power and practical interpretability so that it can be used by medical consultation platforms and service providers.

The high performance across the ML algorithms demonstrates that our 11-feature model is a useful predictive tool (research objective 1). In terms of feature importance (research objectives 2 and 3), our results show that although physician reputation is important, service delivery quality and patient involvement appear to contribute more to the payment. We further identified five patient subgroups based on DT feature configurations. The configurations show how features related to patient characteristics, platform functionalities, and patient-provider interaction are combined to result in different payment outcomes. These configurations highlight the offline connection and responsive service delivery as key differentiators for payment vs free trial–only services.

Limitations

First, decisions made during the feature selection procedure may cause bias in the subsequent analysis. Although the results of this study achieved satisfactory overall performance, a different set of features that are comparable with the current ones can be used to cross-validate our model.

Second, although the platform provides various long- and short-term service options, to ensure consistency in data cleaning and interpretability of results, we only included short-term services based on the service tags available. However, future research should examine long-term service subscription, as patients' decision-making criteria can be very different than for short-term service subscription.

Third, considering problems with data quality and limited variability, we did not include the platform's online physician reputation ratings. However, future research could focus on physicians whose ratings do vary over time to observe how noticeable changes in ratings influence payment.

Fourth, our analysis was based on the Chinese context. Considering the cultural differences and health care regulations, our results may have limited generalizability to other contexts. However, the mechanisms and types of interactions that have been found are generic enough to be promoted and managed in different online medical consultation platforms and in different countries. Furthermore, the Chinese context itself is quite large and should be of interest on its own.

Comparison With Prior Work

Although the majority of features in our predictive model were examined in existing research on payment for online medication consultation, several new features specific to this type of platform and some surprising differences from existing research also emerged. Unexpectedly, physicians' offline reputation, as indicated by the title and the affiliated hospital ranking, does not rank high in the ML algorithms and does not appear in the top three levels of the tree structure. These physician offline reputation features are frequently employed by previous studies in similar contexts [7,15]. Although our LR results exhibit significant coefficients for these offline reputation features, in the tree structure, they only play a role in combination with other features in the lower levels. It is likely that patients experience different stages of awareness and learning during the phases of physician selection, free service, and paid service [9,34]. Although physician reputation may increase patients' initial service awareness and influence physician selection, it seems that service experience (ie, service quality and intensive involvement) is a more important payment differentiator. Thus, our results show that regression may not be the best method to detect the impacts of various predictors and may yield oversimplified interpretation—regression only shows a linear additive relationship and excludes collinearity, whereas in reality, complex interactions and multiple paths to payment may exist.

In contrast to previous results that show the positive influence of prior physician-patient social ties on payment [18], our results show that a prior offline relationship with the physician does not always seem to be a facilitating factor for online payment. Although one subgroup of offline patients with existing social ties with the physician exhibits interactive service experiences and makes online payments, another offline subgroup seems to only use free services for simple follow-ups without deepening the online portion of the relationship and thus avoiding payment. Thus, it may be difficult for patients to completely shift their health care practices and habits from the offline to the online setting.

Previous studies also highlight virtual gifts as a positive signal for payment [2,12]. However, our findings suggest that virtual gifts may be a double-edged sword. For patients who have no prior offline connections with the physician, allowing them to show gratitude with a virtual gift function may not be a good strategy, as this type of patient may substitute this virtual gift for payment. However, if the service is intensive, virtual gifts and payment will be additive rather than substitutive.

In line with previous literature on online service delivery, responsive service is a key antecedent of payment [35,36]. Encouraging patient engagement (eg, encouraging multiple timely interactions with the physician) may help promote payment. As each response to the physician counts as one free trial for the patients, reluctance to consult further may arise at the end of each conversation turn. Persuading patients to keep on responding in a timely manner should be beneficial for establishing long-term patient-physician collaboration and attracting payments.

Previous studies in similar contexts generally use a linear regression approach; however, we employ ML—with its ability to mine massive fine-grained behavior data [37]—to explore the associations and predictive power of various consultation service–related features. The various classifiers based on different ML philosophies for a binary classification problem provide complementary views of how the model can help us understand payment. The feature ranking and configuration results from four ML approaches indicate that these features are not generating linear impacts, a finding that was not evident in previous studies.

Conclusions

Online delivery of health care services is increasingly common and gives patients a new channel and expanded options for accessing health care services. However, many online medical consultation platforms are struggling to attract and retain patients who are willing to pay, and health care providers on the platform have the additional challenge of standing out in a crowd of physicians who can provide comparable services. This study explores the key features that contribute to patient payment in the online health care consultation market. By mining massive consultation data using ML approaches, our results show that features related to service delivery quality (eg, consultation dialog intensity and physician response rate), patient source (eg, online vs offline returning patients), and patient involvement (eg, provide social returns and reveal previous treatment) appear to contribute more to the patient's payment decision than features related to physician reputation. We further identified five key feature configurations to help classify different interaction patterns between patients and physicians, which result in different payment outcomes.

Acknowledgments

This research was undertaken, in part, thanks to funding from the Canada Research Chairs program (awarded to the second author) and the Natural Science Foundation of China (Nos 71301172, 71571180 awarded to the third author).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Background information.
[DOCX File , 445 KB - medinform_v8i2e16765_app1.docx]

Multimedia Appendix 2
Data cleaning and analysis pipeline.
[DOCX File , 33 KB - medinform_v8i2e16765_app2.docx]

Multimedia Appendix 3
Data-driven feature selection.
[DOCX File , 20 KB - medinform_v8i2e16765_app3.docx]

Multimedia Appendix 4
Additional analysis results.
[DOCX File , 108 KB - medinform_v8i2e16765_app4.docx]

References

1. Li J, Zhang Y, Ma L, Liu X. The impact of the internet on health consultation market concentration: an econometric analysis of secondary data. J Med Internet Res 2016 Oct 28;18(10):e276 [FREE Full text] [doi: 10.2196/jmir.6423] [Medline: 27793793]
2. Yang H, Zhang X. Investigating the effect of paid and free feedback about physicians' telemedicine services on patients' and physicians' behaviors: panel data analysis. J Med Internet Res 2019 Mar 22;21(3):e12156 [FREE Full text] [doi: 10.2196/12156] [Medline: 30900997]
3. Yang Y, Zhang X, Lee PK. Improving the effectiveness of online healthcare platforms: an empirical study with multi-period patient-doctor consultation data. Int J Prod Econ 2019;207:70-80. [doi: 10.1016/j.ijpe.2018.11.009]
4. Cordina J, Jones EP, Kumar R, Martin CP. McKinsey & Company. 2018 Jul. Healthcare Consumerism 2018: An Update on the Journey URL: https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/healthcare-consumerism-2018 [accessed 2019-05-05]
5. Garner Insights. 2018 Mar. Online Doctor Consultation Market: Global Market Synopsis, Growth Factors, Industry Segmentation, Regional Analysis And Competitive Analysis 2017 - 2025 URL: http://garnerinsights.com/

Online-Doctor-Consultation-Market-Global-Market-Synopsis-Growth-Factors-Industry-Segmentation-Regional-Analysis-And-Competitive-Analysis-2017---2025 [accessed 2019-05-05]
6. Cao X, Liu Y, Zhu Z, Hu J, Chen X. Online selection of a physician by patients: empirical study from elaboration likelihood perspective. Comput Human Behav 2017 Aug;73:403-412. [doi: 10.1016/j.chb.2017.03.060]
7. Deng Z, Hong Z, Zhang W, Evans R, Chen Y. The effect of online effort and reputation of physicians on patients' choice: 3-wave data analysis of China's good doctor website. J Med Internet Res 2019 Mar 8;21(3):e10170 [FREE Full text] [doi: 10.2196/10170] [Medline: 30848726]
8. Yu H, Xiang K, Yu J. Understanding a Moderating Effect of Physicians' Endorsement to Online Workload: An Empirical Study in Online Health-Care Communities. In: Proceedings of the 2017 IEEE International Conference on Big Data. 2017 Presented at: Big Data'17; December 11-14, 2017; Boston, MA. [doi: 10.1109/bigdata.2017.8258570]
9. Li J, Tang J, Jiang L, Yen DC, Liu X. Economic success of physicians in the online consultation market: a signaling theory perspective. Int J Electron Comm 2019 Mar;23(2):244-271. [doi: 10.1080/10864415.2018.1564552]
10. Greenhalgh T, Vijayaraghavan S, Wherton J, Shaw S, Byrne E, Campbell-Richards D, et al. Virtual online consultations: advantages and limitations (VOCAL) study. BMJ Open 2016 Jan 29;6(1):e009388 [FREE Full text] [doi: 10.1136/bmjopen-2015-009388] [Medline: 26826147]
11. Singh AP, Joshi HS, Singh A, Agarwal M, Kaur P. Online medical consultation: a review. Int J Community Med Public Health 2018;5(4):1230-1232. [doi: 10.18203/2394-6040.ijcmph20181195]
12. Yang H, Guo X, Wu T, Ju X. Exploring the effects of patient-generated and system-generated information on patients' online search, evaluation and decision. Electron Commer Res Appl 2015;14(3):192-203. [doi: 10.1016/j.elerap.2015.04.001]
13. Kumar V. Making freemium work. Harvard business review. . ISSN 2014;92(5):0017-0019.
14. Liu X, Guo X, Wu H, Wu T. The impact of individual and organizational reputation on physicians' appointments online. Int J Electron Comm 2016 Jun;20(4):551-577. [doi: 10.1080/10864415.2016.1171977]
15. Guo S, Guo X, Fang Y, Vogel D. How doctors gain social and economic returns in online health-care communities: a professional capital perspective. J Manag Inf Syst 2017 Aug;34(2):487-519. [doi: 10.1080/07421222.2017.1334480]
16. Liu J, Bian Y, Ye Q, Jing D. Free for Caring? The effect of offering free online medical-consulting services on physician performance in e-health care. Telemed J E Health 2019 Oct;25(10):979-986. [doi: 10.1089/tmj.2018.0216] [Medline: 30566383]
17. Lu N, Wu H. Exploring the impact of word-of-mouth about physicians' service quality on patient choice based on online health communities. BMC Med Inform Decis Mak 2016 Nov 26;16(1):151 [FREE Full text] [doi: 10.1186/s12911-016-0386-0] [Medline: 27888834]
18. Guo S, Guo X, Zhang X, Vogel D. Doctor–patient relationship strength's impact in an online healthcare community. Inform Technol Dev 2017 Mar;24(2):279-300. [doi: 10.1080/02681102.2017.1283287]
19. Li Y, Ma X, Song J, Yang Y, Ju X. Exploring the effects of online rating and the activeness of physicians on the number of patients in an online health community. Telemed J E Health 2019 Nov;25(11):1090-1098. [doi: 10.1089/tmj.2018.0192] [Medline: 30676279]
20. Hu MM, Yang S, Xu DY. Understanding the social learning effect in contagious switching behavior. Manage Sci 2019;65(10):4771-4794. [doi: 10.1287/mnsc.2018.3173]
21. Jiang Y, Ho YC, Yan X, Tan Y. Investor platform choice: herding, platform attributes, and regulations. J Manag Inf Syst 2018 Mar;35(1):86-116. [doi: 10.1080/07421222.2018.1440770]
22. Tsai HT, Bagozzi RP. Contribution behavior in virtual communities: cognitive, emotional, and social influences. MIS Q 2014;38(1):143-163. [doi: 10.25300/misq/2014/38.1.07]
23. Luo P, Chen K, Wu C, Li Y. Exploring the social influence of multichannel access in an online health community. J Assoc Inf Sci Technol 2018;69(1):98-109. [doi: 10.1002/asi.23928]
24. Daemmrich A. The political economy of healthcare reform in China: negotiating public and private. Springerplus 2013;2:448 [FREE Full text] [doi: 10.1186/2193-1801-2-448] [Medline: 24052932]
25. Yu L, Liu H. Feature Selection for High-dimensional Data: A Fast Correlation-Based Filter Solution. In: Proceedings of the 20th International Conference on Machine Learning. 2003 Presented at: ICML'03; August 21-24, 2003; Washington, DC.
26. KNIME. 2015. Seven Techniques for Data Dimensionality Reduction URL: https://www.knime.com/blog/seven-techniques-for-data-dimensionality-reduction [accessed 2019-05-01]
27. Hall MA. Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. In: Proceedings of the Seventeenth International Conference on Machine Learning. 2000 Presented at: ICML'00; June 29 - July 2, 2000; CA, USA.
28. Batista GE, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor Newsl 2004;6(1):20. [doi: 10.1145/1007730.1007735]
29. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag 2009;45(4):427-437. [doi: 10.1016/j.ipm.2009.03.002]
30. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat 2001;29(5):1189-1232. [doi: 10.1214/aos/1013203451]

31. Bäckman L, Dixon RA. Psychological compensation: a theoretical framework. Psychol Bull 1992 Sep;112(2):259-283. [doi: 10.1037/0033-2909.112.2.259] [Medline: 1454895]
32. Angst CM, Agarwal R. Adoption of electronic health records in the presence of privacy concerns: the elaboration likelihood model and individual persuasion. MIS Q 2009 Jun;33(2):339-370. [doi: 10.2307/20650295]
33. Bansal G, Zahedi F, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decis Support Syst 2010 May;49(2):138-150. [doi: 10.1016/j.dss.2010.01.010]
34. Liu CZ, Au YA, Choi HS. Effects of freemium strategy in the mobile app market: an empirical study of google play. J Manag Inf Syst 2014;31(3):326-354. [doi: 10.1080/07421222.2014.995564]
35. Storey C, Cankurtaran P, Papastathopoulou P, Hultink EJ. Success factors for service innovation: a meta-analysis. J Prod Innov Manag 2016 Sep;33(5):527-548. [doi: 10.1111/jpim.12307]
36. Wakefield KL, Blodgett JG. Customer response to intangible and tangible service factors. Psychol Mark 1999 Jan;16(1):51-68. [doi: 10.1002/(sici)1520-6793(199901)16:1<51::aid-mar4>3.0.co;2-0]
37. Martens D, Provost F, Clark J, de Fortuny EJ. Mining massive fine-grained behavior data to improve predictive analytics. MIS Q 2016;40(4):869-888. [doi: 10.25300/misq/2016/40.4.04]

Abbreviations

AUC: area under the receiver operating characteristic curve
DT: decision tree
GB: gradient boost
LR: logistic regression
ML: machine learning
RF: random forest

Edited by G Eysenbach; submitted 22.10.19; peer-reviewed by M Sokolova, A Benis, WD Dotson, R Segall; comments to author 19.11.19; revised version received 14.01.20; accepted 24.01.20; published 18.02.20.

Please cite as:
Jiang J, Cameron AF, Yang M
Analysis of Massive Online Medical Consultation Service Data to Understand Physicians' Economic Return: Observational Data Mining Study
JMIR Med Inform 2020;8(2):e16765
URL: http://medinform.jmir.org/2020/2/e16765/
doi:10.2196/16765
PMID:32069213

Original Paper

Just Because (Most) Hospitals Are Publishing Charges Does Not Mean Prices Are More Transparent

Cody Lendon Mullens[1,2*], MPH; J Andres Hernandez[3,4*], BS; Evan D Anderson[5*], JD, PhD; Lindsay Allen[6*], MA, PhD

[1]West Virginia University School of Medicine, Morgantown, WV, United States

[2]Center for Public Health Initiatives, University of Pennsylvania, Philadelphia, PA, United States

[3]The Wharton School, University of Pennsylvania, Philadelphia, PA, United States

[4]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[5]Center for Public Health Initiatives, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[6]Department of Health Policy, Management, and Leadership, School of Public Health, West Virginia University, Morgantown, WV, United States

[*]all authors contributed equally

Corresponding Author:
Cody Lendon Mullens, MPH
Center for Public Health Initiatives
University of Pennsylvania
3620 Hamilton Walk
Philadelphia, PA,
United States
Phone: 1 3047238039
Email: cmullen3@mix.wvu.edu

Abstract

Background: The Centers for Medicare and Medicaid Services (CMS) recently mandated that all hospitals publish their charge description masters (CDMs) online, in a machine-readable format, by January 1, 2019. In addition, CMS recommended that CDM data be made available in a manner that was consumer friendly and accessible to patients.

Objective: This study aimed to (1) examine all hospitals across the state of Pennsylvania to understand policy compliance and (2) use established metrics to measure accessibility and consumer friendliness of posted CDM data.

Methods: A cross-sectional analysis was conducted to quantify hospital website compliance with the recent CMS policies requiring hospitals to publish their CDM. Data were collected from all Pennsylvania hospital websites. Consumer friendliness was assessed based on searchability, number of website clicks to data, and supplemental educational materials accompanying CDMs such as videos or text.

Results: Most hospitals (189/234, 80.1%) were compliant, but significant variation in data presentation was observed. The mean number of website clicks to the CDM was 3.7 (SD 1.3; range: 1-8). A total of 23.1% of compliant hospitals provided no supplemental educational material with their CDM.

Conclusions: Although disclosure of charges has improved, the data may not be sufficient to meaningfully influence patient decision making.

(JMIR Med Inform 2020;8(2):e14436) doi:10.2196/14436

KEYWORDS

health care costs; delivery of health care; health policy

Introduction

As of 2017, national health care expenditures in the United States rose to US $3.5 trillion, an increase of close to 4% compared with the previous year [1]. With a stated objective of empowering patients and reducing administrative burdens, the Centers for Medicare and Medicaid Services (CMS) mandated that all hospitals publish charge description master (CDM) data online [2,3]. A hospital CDM is a comprehensive list of a hospital's charges to patients or health insurance companies for services rendered during a hospital stay. One rationale for the policy is that increased price transparency will encourage

patients to *shop around* for competitively priced health care services, much as they would for a new car [4-6].

Recent changes in Medicare's payment policies under the inpatient prospective payment system (PPS) and the long-term care hospital PPS required that the CDM be made available in a machine-readable format by January 1, 2019 [2]. Only Veterans Affairs (VA) hospitals and hospitals reimbursed under state or local cost control systems are exempt from the requirements. Machine-readable format refers to documents that are digitally accessible and in file formats that are easily processed by computers (ie, comma separated value [CSV] or XML files). Notably, a "frequently asked questions" form published by CMS recommended that CDM data be published in a consumer-friendly, accessible manner, which is an important consideration given that the target consumers of the data are patients [7]. Given the inherent confusion surrounding CDMs—including their relationship to actual prices—even the savviest consumers are unlikely to know how to interpret them [8]. The purpose of this analysis was to assess hospital compliance with the CDM policy and determine accessibility and consumer friendliness of CDM data for all hospitals within the state of Pennsylvania.

Methods

Study Design

We examined the presentation of CDM data for all hospitals in the state of Pennsylvania. This state was chosen because of its relatively large number of hospitals and for its variation in rural versus urban health systems, facility types, and hospital tax status. All hospitals were identified using data from the Hospital and Healthsystem Association of Pennsylvania and were categorized as nonprofit, for profit, city, state, or federal based on tax status. In addition, hospitals were classified into the following facility types: general acute, general acute specialty, rehabilitation, psychiatric, VA, long-term acute care, drug/alcohol, maternity, and other specialty hospitals.

Quantifying Compliance and Consumer Friendliness

Each hospital website was accessed on January 7, 2019, and queried for online CDM publication in a machine-readable format, which was required for compliance with the policy. Hospitals that were noncompliant at initial analysis were reassessed 1 week later on January 14, 2019. Accessibility of the CDMs was determined based on the number of page clicks required to access the CDM from the hospital home page, an established method for assessing website usability and navigability [9,10]. Accessibility of CDMs was further assessed through utilization of hospital website's search function, using keywords such as "chargemaster," "charge master," "charge description master," "standard charges," "charge," and "price."

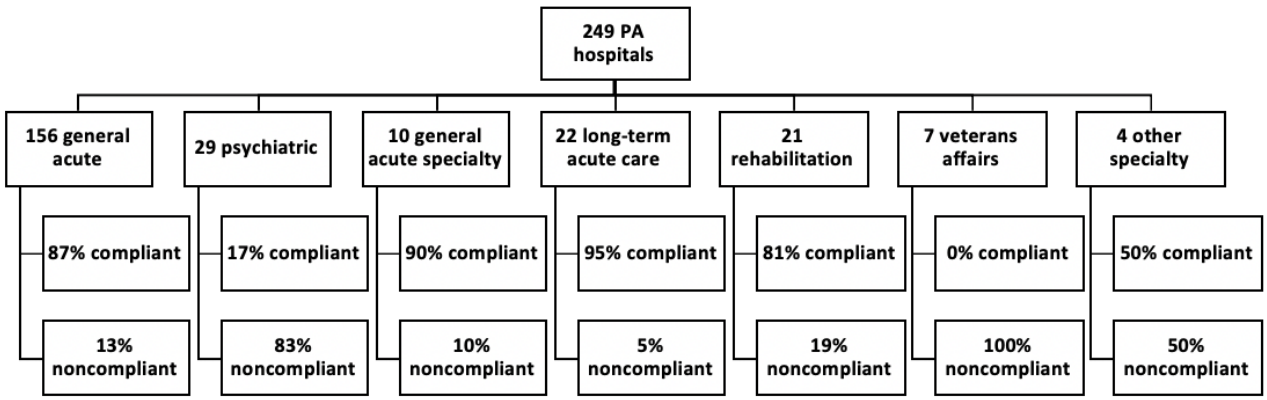
To measure consumer friendliness, we next determined if any supplemental information was provided by hospitals to help patients understand the provided materials such as descriptions as to what CDMs are, their utility, or how patients can interpret them in the context of price of care. Finally, all hospital home pages were evaluated based on their display of CDMs or other financial information for viewers and prospective patients. Descriptive statistics were calculated in Stata (StataCorp LLC, version IC 15.1).

Results

Hospital Demographics

A total of 249 hospitals were identified and included in the study (Figure 1). Most hospitals were nonprofit tax status (148/249, 59.4%), followed by for profit (86/249, 34.5%) and federal/city/state (15/249, 6.1%). The most common hospital type was found to be general acute care hospitals (156/249, 62.7%), followed by psychiatric (29/249, 11.6%), long-term acute care (22/249, 8.8%), rehabilitation (21/249, 8.4%), general acute specialty hospitals (10/249, 4.0%), VA (7/249, 2.8%), and other specialty hospitals (4/249, 1.6%).

Figure 1. Demographic breakdown of Pennsylvania hospitals and compliance with charge description master policy.



Hospital Compliance With Policy

Most hospitals included CDMs in their website (189/249, 74.9% hospitals). Excluding VA hospitals and hospitals reimbursed under state or local cost control systems, which are excluded from the policy, compliance rose to 80.1% (189/ 234) hospitals.

A total of 20 hospitals were out of compliance with the mandated policy because they did not comply with the mandated formatting of CDM data. Moreover, 10 hospitals posted online databases, 1 posted online text, and 9 posted PDFs, none of which were in a machine-readable format as required.

Analyzing Accessibility and Consumer Friendliness

Among the 189 compliant hospitals, 116 posted billing, pricing, or other financial information (such as payment plans or hospital financial resources) on the hospital home page. Few hospitals included a direct hyperlink or information regarding CDM data and the recent mandate on their hospital website's home page (21 out of the 189 facilities posting data). Although it was possible to access the CDM through link-clicking on most of the compliant websites, 21 hospitals required users to utilize the search function within the hospital website to obtain their CDM. In addition to the 21 hospitals who only permitted access to CDM data through the search function, an additional 126 hospitals enabled users to query searches within the hospital website to access published CDMs.

Mean number of clicks to CDM access was 3.7 (SD 1.3; range: 1-8 clicks; Figure 2). Hospital CDM data end points were predominantly linked and downloadable CSV or XML (files for Microsoft Excel) files (156 hospitals; Table 1).

Finally, we assessed the prevalence of supplemental financial information for users and prospective patients regarding CDMs, cost, charges, and any additional relevant information. Although most hospitals provided supplemental text information on the website for viewers (107/190, ie, 56.3% of compliant hospitals), a substantial proportion (23.1%) did not provide any information. Less commonly, hospitals included text and video information (21/249, 8.4%) or video only (2.1%). Interestingly, 7.9% of hospitals included disclaimers to users or required acknowledgment alluding to the insufficiency of CDMs for determining actual price of care within their text and other supplemental information.

Figure 2. Distribution of hospital website charge description master access based on the number of clicks to download charge description master data. CDM: charge description master.

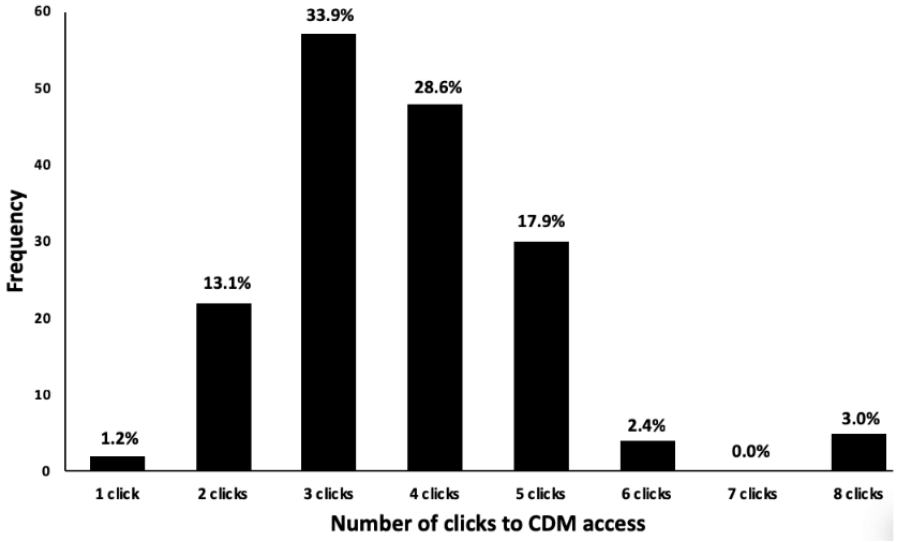


Table 1. Hospital website charge description master end points.

Website end point	Hospitals, n (%)
Interactive online database[a]	10 (5.3)
Downloadable comma separated value or XML file	169 (89.4)
Downloadable PDF[a]	9 (4.8)
Online text[a]	1 (0.5)

[a]Interactive online databases, PDFs, and online text are not technically compliant with the stated policy.

Discussion

Principal Findings

This is the first known study to quantify compliance with the new CMS CDM transparency requirement. Although 80.1% (189/234) of hospitals in Pennsylvania are compliant with the policy by posting CDM data at the time of our study, wide variation was observed in the degree of accessibility of the CDM and specific compliance with mandated formatting.

One-fifth of Pennsylvania hospitals were noncompliant with the policy, which is not unprecedented; in California, hospitals have been reluctant to comply with regulations mandating that hospitals provide uninsured patients with price estimates to promote health service price shopping [11]. Even with full compliance, however, there are reasons to question the value of the requirements in helping slow the growth of health care costs.

First, consumers of health care struggle to interpret nomenclature around health care financing such as understanding and differentiating between charge and price [12]. A charge is the

dollar amount associated with a particular medical service before payer discount negotiation, whereas a price is the negotiated and contracted dollar amount for the payer. In Pennsylvania, many hospitals did not provide accompanying resources with CDM data to help the public understand these differences. Amendments to the policy that require such accompanying resources with CDM data may be helpful and informative to those patients who do access these data.

Second, when consumers are covered by health insurance plans, they are not responsible for negotiating prices of services rendered or bearing the full cost of care. This leads to the dilemma of moral hazard, wherein patients use more medical care than they would had they not been covered by insurance. In the context of health care, providers (ie, hospitals) may feel less obligated to be transparent regarding charges to consumers, realizing that most consumers never bear the full cost of care or the charge associated with CDM. Recent attempts at increasing price transparency to encourage "shopping" in health care have produced mixed results. Sinaiko and Rosenthal [13] studied an insured, nonelderly adult population's use of a payer-developed payment estimator; they found considerable engagement with and utilization of the estimator, especially among those who were younger, with fewer comorbid conditions, and with relatively high health care system utilization rates. The authors concluded that tools to increase transparency of price in health care have the potential to meaningfully impact patient decision making and health care service utilization [13]. Other studies have failed to find significant changes in actual health care spending associated with implementation and availability of different transparency tools [14,15]. Americans attest to and support the utility of tools for assisting in health care price shopping, but few patients actually seek out health care price–related information in practice [16].

In contrast to consumers who are insulated from much of the costs, many individuals have high deductible health plans, which are consumer-driven in the sense that they bear a much more substantial cost burden when seeking care. This specific population may stand to benefit from published CDM data as a means of shopping for health care services. However, a considerable portion of this population may also be effectively incapable of paying almost any out-of-pocket costs. As a Federal Reserve study recently noted, 40% of Americans do not have the accessible assets to pay a US $400 emergency expense [17]. This underscores both how unpredictable bills can destabilize households and how precarity in household assets can destabilize theories that consumers will vigorously shop for health care services for based on price.

Finally, lack of accessibility and/or consumer friendliness may impede prospective patients from using the CDMs. For example, some CDM data could only be accessed through a focused query using the website's search function, which patients may not know to use. Other hospitals placed links to the data within unrelated or unlabeled subsections of the website. Moreover, CDM data are generally written in medical jargon likely indecipherable by the general public, using incomprehensive acronyms or technical names of procedures and equipment. This is an important barrier to consider not only for those who have lower health care literacy but also for those who are less adept with computer- and internet-related technologies.

Future Directions and Limitations

Although the CMS policy is a step in the right direction for increased transparency, the consumer friendliness of these data and the direct implications on patients or their health insurance provider are unclear. One clear benefit of the charge data availability, however, is that researchers will be able to study the significant charge variability that exists between comparable health care facilities and hospitals. Future policies may consider additional steps toward true price transparency for both services rendered as well as pharmaceuticals and medical devices.

Limitations of this study include the single-state analysis; though as mentioned earlier, Pennsylvania was chosen because of its diversity in geography, demographics, and facility types. In addition, because of the cross-sectional study design, it is impossible to make any inferences about the causal relationship between the policy changes and observations about the accessibility of CDM data in Pennsylvania hospitals.

Conclusions

The majority of hospitals in Pennsylvania have complied with the CDM policy. However, there is considerable variation in the accessibility and consumer friendliness of the CDMs. Determining whether enhanced access to CDM data will alter consumer or institutional behavior remains an important priority for future health services research.

Conflicts of Interest

None declared.

References

1. Centers for Medicare and Medicaid Services. Historical URL: https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nationalhealthaccountshistorical.html [accessed 2019-12-19]
2. Centers for Medicare and Medicaid Services. Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care Hospital Prospective Payment System and Proposed Policy Changes and Fiscal Year 2019 Rates; Proposed Quality Reporting Requirements for Specific Providers; Proposed Medicare and Medicaid Electronic Health Record (EHR) Incentive Programs (Promoting Interoperability Programs) Requirements for Eligible Hospitals, Critical Access Hospitals, and Eligible Professionals; Medicare Cost Reporting Requirements; and Physician Certification and Recertification of Claims. CMS-1694-P URL: https://www.cms.gov/Medicare/

Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2019-IPPS-Proposed-Rule-Home-Page-Items/FY2019-IPPS-Proposed-Rule-Regulations [accessed 2019-12-19]
3. Centers for Medicare and Medicaid Services. 2018 Aug 2. CMS Finalizes Changes to Empower Patients and Reduce Administrative Burden [press release] URL: https://www.cms.gov/newsroom/press-releases/cms-finalizes-changes-empower-patients-and-reduce-administrative-burden [accessed 2019-12-19]
4. Ginsburg PB. Shopping for price in medical care. Health Aff (Millwood) 2007;26(2):w208-w216. [doi: 10.1377/hlthaff.26.2.w208] [Medline: 17284467]
5. Bloche MG. Consumer-directed health care. N Engl J Med 2006;355(17):1756-1759. [doi: 10.1056/nejmp068127]
6. Sinaiko AD, Rosenthal MB. Increased price transparency in health care--challenges and potential effects. N Engl J Med 2011 Mar 10;364(10):891-894. [doi: 10.1056/NEJMp1100041] [Medline: 21388306]
7. Centers for Medicare and Medicaid Services. Additional Frequently Asked Questions Regarding Requirements for Hospitals To Make Public a List of Their Standard Charges via the Internet URL: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ProspMedicareFeeSvcPmtGen/Downloads/Additional-Frequently-Asked-Questions-Regarding-Requirements-for-Hospitals-To-Make-Public-a-List-of-Their-Standard-Charges-via-the-Internet.pdf [accessed 2019-12-19]
8. Appleby J, Ostrov BF. Kaiser Health News. 2019 Jan 4. As Hospitals Post Sticker Prices Online, Patients Will Remain Befuddled URL: https://khn.org/news/as-hospitals-post-sticker-prices-online-most-patients-will-remain-befuddled/ [accessed 2019-12-19]
9. Waisberg D, Kaushik A. Web Analytics 2.0: Empowering Customer Centricity. Search Engine Mark J 2009;2(1):5-11 [FREE Full text]
10. Farney TA. Click analytics: visualizing website use data. Inf Technol Librar 2011;30(3):141-148. [doi: 10.6017/ital.v30i3.1771]
11. Farrell KS, Finocchio LJ, Trivedi AN, Mehrotra A. Does price transparency legislation allow the uninsured to shop for care? J Gen Intern Med 2010 Feb;25(2):110-114 [FREE Full text] [doi: 10.1007/s11606-009-1176-5] [Medline: 19936845]
12. Huston SJ. Measuring financial literacy. J Consum Aff 2010;44(2):296-316. [doi: 10.2139/ssrn.1945216]
13. Sinaiko AD, Rosenthal MB. Examining a health care price transparency tool: who uses it, and how they shop for care. Health Aff (Millwood) 2016 Apr;35(4):662-670. [doi: 10.1377/hlthaff.2015.0746] [Medline: 27044967]
14. Desai S, Hatfield LA, Hicks AL, Sinaiko AD, Chernew ME, Cowling D, et al. Offering a price transparency tool did not reduce overall spending among California public employees and retirees. Health Aff (Millwood) 2017 Aug 1;36(8):1401-1407. [doi: 10.1377/hlthaff.2016.1636] [Medline: 28784732]
15. Desai S, Hatfield LA, Hicks AL, Chernew ME, Mehrotra A. Association between availability of a price transparency tool and outpatient spending. J Am Med Assoc 2016 May 3;315(17):1874-1881. [doi: 10.1001/jama.2016.4288] [Medline: 27139060]
16. Mehrotra A, Dean KM, Sinaiko AD, Sood N. Americans support price shopping for health care, but few actually seek out price information. Health Aff (Millwood) 2017 Aug 1;36(8):1392-1400. [doi: 10.1377/hlthaff.2016.1471] [Medline: 28784731]
17. Federal Reserve Board. 2018 May. Report on the Economic Well-Being of US Households in 2017 URL: https://www.federalreserve.gov/publications/files/2017-report-economic-well-being-us-households-201805.pdf [accessed 2019-12-19]

Abbreviations

CDM: charge description master
CMS: Centers for Medicare and Medicaid Services
CSV: comma separated value
PPS: prospective payment system
VA: Veterans Affairs

Edited by G Eysenbach; submitted 18.04.19; peer-reviewed by K Patel, J Colquitt, Jr; comments to author 30.07.19; revised version received 30.08.19; accepted 22.10.19; published 06.02.20.

Please cite as:
Mullens CL, Hernandez JA, Anderson ED, Allen L
Just Because (Most) Hospitals Are Publishing Charges Does Not Mean Prices Are More Transparent
JMIR Med Inform 2020;8(2):e14436
URL: http://medinform.jmir.org/2020/2/e14436/
doi:10.2196/14436
PMID:

Original Paper

Expedited Safety Reporting Through an Alert System for Clinical Trial Management at an Academic Medical Center: Retrospective Design Study

Yu Rang Park[1], PhD; HaYeong Koo[2], MS; Young-Kwang Yoon[2], PhD; Sumi Park[3], BS; Young-Suk Lim[3,4], MD, PhD; Seunghee Baek[5], PhD; Hae Reong Kim[1], MS; Tae Won Kim[2,6], MD, PhD

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Clinical Research Center, Asan Institute of Life Sciences, Asan Medical Center, Seoul, Republic of Korea

[3]Clinical Trial Center, Asan Medical Center, Seoul, Republic of Korea

[4]Department of Gastroenterology, Liver Center, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[5]Department of Clinical Epidemiology and Biostatistics, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[6]Department of Oncology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Tae Won Kim, MD, PhD
Clinical Research Center
Asan Institute of Life Sciences
Asan Medical Center
88, Olympic-ro 43-gil, Songpa-gu
Seoul, 05505
Republic of Korea
Phone: 82 2 3010 3910
Fax: 82 2 3010 3910
Email: twkimmd@amc.seoul.kr

Abstract

Background: Early detection or notification of adverse event (AE) occurrences during clinical trials is essential to ensure patient safety. Clinical trials take advantage of innovative strategies, clinical designs, and state-of-the-art technologies to evaluate efficacy and safety, however, early awareness of AE occurrences by investigators still needs to be systematically improved.

Objective: This study aimed to build a system to promptly inform investigators when clinical trial participants make unscheduled visits to the emergency room or other departments within the hospital.

Methods: We developed the Adverse Event Awareness System (AEAS), which promptly informs investigators and study coordinators of AE occurrences by automatically sending text messages when study participants make unscheduled visits to the emergency department or other clinics at our center. We established the AEAS in July 2015 in the clinical trial management system. We compared the AE reporting timeline data of 305 AE occurrences from 74 clinical trials between the preinitiative period (December 2014-June 2015) and the postinitiative period (July 2015-June 2016) in terms of three AE awareness performance indicators: onset to awareness, awareness to reporting, and onset to reporting.

Results: A total of 305 initial AE reports from 74 clinical trials were included. All three AE awareness performance indicators were significantly lower in the postinitiative period. Specifically, the onset-to-reporting times were significantly shorter in the postinitiative period (median 1 day [IQR 0-1], mean rank 140.04 [SD 75.35]) than in the preinitiative period (median 1 day [IQR 0-4], mean rank 173.82 [SD 91.07], $P \leq .001$). In the phase subgroup analysis, the awareness-to-reporting and onset-to-reporting indicators of phase 1 studies were significantly lower in the postinitiative than in the preinitiative period (preinitiative: median 1 day, mean rank of awareness to reporting 47.94, vs postinitiative: median 0 days, mean rank of awareness to reporting 35.75, $P=.01$; and preinitiative: median 1 day, mean rank of onset to reporting 47.4, vs postinitiative: median 1 day, mean rank of onset to reporting 35.99, $P=.03$). The risk-level subgroup analysis found that the onset-to-reporting time for low- and high-risk studies significantly decreased postinitiative (preinitiative: median 4 days, mean rank of low-risk studies 18.73, vs postinitiative: median 1 day, mean rank of low-risk studies 11.76, $P=.02$; and preinitiative: median 1 day, mean rank of high-risk studies 117.36, vs postinitiative: median 1 day, mean rank of high-risk studies 97.27, $P=.01$). In particular, onset to reporting was reduced more in the low-risk trial than in the high-risk trial (low-risk: median 4-0 days, vs high-risk: median 1-1 day).

Conclusions: We demonstrated that a real-time automatic alert system can effectively improve safety reporting timelines. The improvements were prominent in phase 1 and in low- and high-risk clinical trials. These findings suggest that an information technology-driven automatic alert system effectively improves safety reporting timelines, which may enhance patient safety.

(JMIR Med Inform 2020;8(2):e14379) doi:10.2196/14379

KEYWORDS

clinical trial; adverse event; early detection; patient safety

Introduction

In recent trends of drug development, proof of concept is rapidly achieved at a low cost, allowing successful projects to be promptly positioned for late-stage development [1,2]; thus, it is critical that investigators be notified of the early signs of a drug's efficacy and safety in real time during clinical trials. The solutions to the challenges of postmarketing evaluation of drug safety require highly collaborative interactions among the US Food and Drug Administration, industry, and other health authorities, as well as the development of registries for spontaneous reporting and epidemiological statistical methodology to detect and interpret adverse event (AE) signals [3-6]. Although clinical trials have rigorous procedures for reporting AEs, few studies have investigated how to detect and report them [7-9].

For reporting, the Consolidated Standards of Reporting Trials (CONSORT) Group generated recommendations regarding the appropriate reporting of AEs [8,9]. With advances in information technology (IT), the reporting of safety and other clinical trial data is moving from paper-based to electronic formats. Typically, electronic reporting portals developed by the clinical trial sponsors and contract research organizations are used as centralized electronic repositories for ongoing clinical trial information to reduce time and cost associated with recordkeeping.

To ensure the safety of study participants, it is crucial to promptly manage AEs and serious adverse events (SAEs), especially during high-risk or early-phase clinical trials. AE management consists of detection, processing, and reporting. AEs in clinical trials are usually detected during scheduled visits or when participants inform investigators of unscheduled visits to the emergency department or clinic. Unscheduled visits are sometimes detected by coordinators during the review of patients' electronic health records. Delayed awareness of AEs among study personnel may jeopardize patient safety, so the prompt detection of unscheduled visits is important during clinical trials. While many health care practitioners understand the importance of AE awareness and reporting in the clinical trial field, they face practical hurdles in systematically managing AEs.

To build a systematic AE management process that addresses issues from occurrence to awareness and reporting, IT support with clinical trial management systems (CTMSs) and electronic medical records (EMRs) can be useful. In July 2015, we established an alert system called the *Adverse Event Awareness System* (AEAS), which was derived from CTMSs and EMRs [10,11]. The AEAS promptly informs investigators when clinical trial participants make unscheduled visits to the emergency room or other departments within the hospital. Such notifications were designed to improve the timelines of AE reporting to stakeholders, such as sponsors, institutional review boards (IRBs), or regulatory bodies. This study investigated the effectiveness of the AEAS by analyzing and comparing the relevant AE reporting timelines before and after the establishment of the AEAS.

Methods

Study Setting

This study was conducted at Asan Medical Center (AMC), a tertiary hospital in Seoul, South Korea. AMC is the largest medical center in Korea, with approximately 2700 inpatient beds and 10,000 outpatient visits per day. AMC has been fully accredited by the Association for the Accreditation of Human Research Protection Program since 2013; the IRB at AMC has received accreditation from the Forum for Ethical Review Committees in the Asian and Western Pacific Region since 2006.

In the process of clinical trials, AMC follows the US Food and Drug Administration regulations, as well as the Korean Good Clinical Practice and International Conference on Harmonization guidelines for AE reporting, which stipulate that all SAEs must be immediately reported to the sponsor except for those that the protocol or Investigator's Brochure identifies as not requiring immediate reporting [12,13]. Sponsors are also required to report fatal and life-threatening *suspected unexpected serious reactions* (SUSARs) to regulatory agencies within 7 days after the sponsor's initial receipt of the information, while other SUSARs may be reported within 15 days.

To manage clinical trials, we implemented a site-specific CTMS in December 2014 [10]. A CTMS is a system for managing clinical trials and research data. The requirements of various organizations related to clinical trials were developed for 14 months and implemented for 12 months based on the Java-based Spring Framework 4.0 (Pivotal Software). The CTMS was developed as an all-in-one system that links hospital information systems, the electronic IRB, enterprise resource planning, and biomaterial management systems in the respective academic medical center. The AEAS was implemented in July 2015 as a submodule of the CTMS interfaced with the EMR to facilitate early awareness of AE occurrences by investigators and study coordinators. The detailed operation process is as follows: if a patient makes an unscheduled visit to the hospital, the EMR confirms that the patient is in a clinical trial; if the patient participates in a specific clinical trial, the patient's information is sent to the CTMS through the application programming

interface; the CTMS then sends a text notification to the principal investigator or clinical research coordinator regarding the patient's emergency room visit or hospitalization.

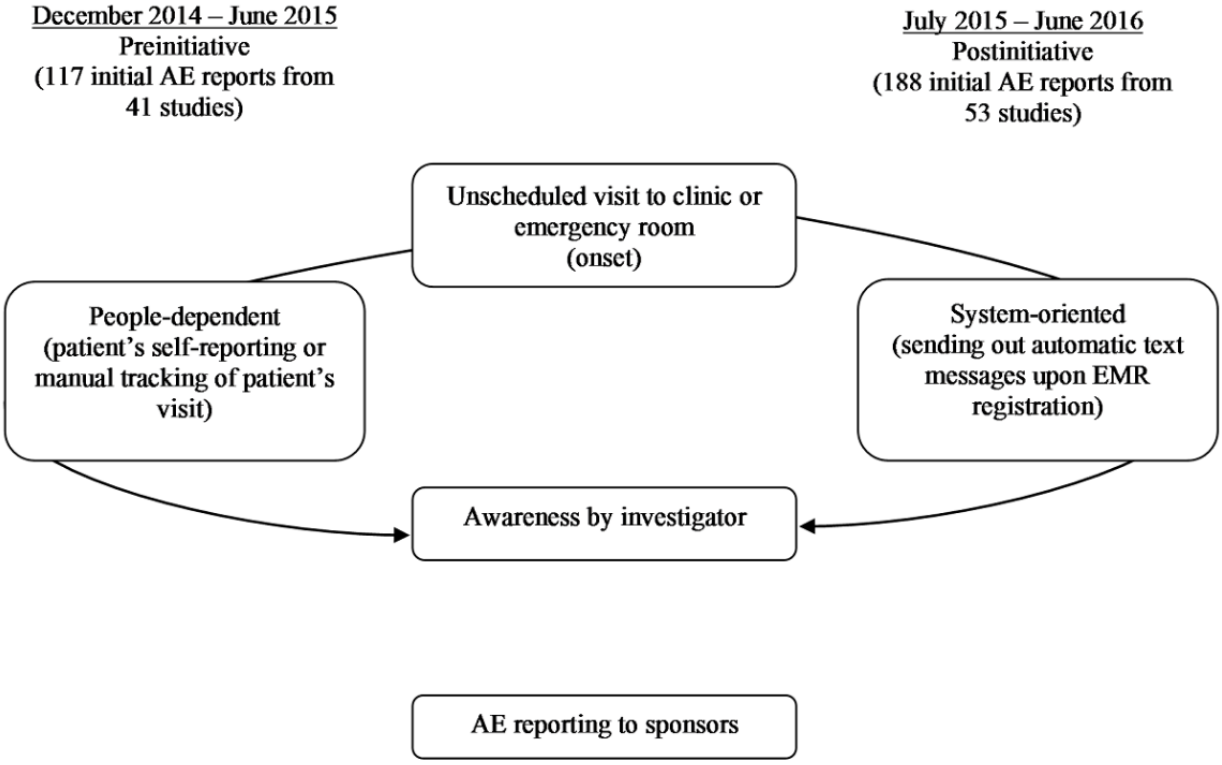
Selection of Clinical Trials

From December 2014 to June 2016, 196 clinical trials were managed through the AMC CTMS. Among these trials, 93 had at least one AE report, with a total of 305 initial, 403 follow-up, and 179 finalized AE reports. In this study, we included 305 initial AE reports from 74 clinical trials to measure the improvement in AE awareness by the implementation of the AEAS. The follow-up or finalized reports were excluded from this study because they could not serve to verify the effectiveness of the AEAS in matters such as early recognition of AEs.

A total of 117 initial AEs were processed in the CTMS prior to the implementation of the AEAS—December 2014 to June 2015—and 188 initial AEs were processed postimplementation—July 2015 to June 2016 (see Figure 1). Overall, patient AE monitoring processes were the same in the pre- and postinitiative periods, but the AE awareness method was different: people-dependent versus system-supported. Of the 74 clinical trials that were evaluated, 41 were carried out during the preinitiative period and 53 during the postinitiative period, with 20 trials overlapping the two periods.

The following data were collected: the dates of initial AE onset when investigators became aware of a given AE and when the AE report was submitted. The intervals between each step were calculated, with AE onset defined as the time when patients made an unscheduled visit to the emergency room or clinics.

Figure 1. Initial adverse event (AE) reports before and after implementation of the Adverse Event Awareness System (AEAS). CTMS: clinical trial management system; EMR: electronic medical record.



Statistical Analysis

General characteristics were compared using chi-square tests. The numbers of days between AE reporting phases (ie, onset, awareness, and reporting) were presented as medians with IQRs and mean ranks with SDs among the periods before and after AEAS implementation. The Shapiro-Wilk normality test showed the time intervals between AE onset, awareness, and reporting to have nonnormal distributions. We, therefore, used the Wilcoxon rank-sum test for comparing the pre- and postinitiative periods. We also presented the effect size of a pre- and postcomparison with a Cohen *d* test. According to US Food and Drug Administration guidelines, AEs are reported as being on business days [14-16]; likewise, we set the intervals using business days. We also performed subgroup analyses according to risk level and study phase. The risk level of each clinical trial was determined based on the classification by the ADAPted MONitoring (ADAMON) project [17]. A higher risk level value means a subject has a higher than lower risk level. *P* values lower than .05 were deemed statistically significant. All analyses were performed in R, version 3.5.3 (The R Foundation).

Ethics Statement

This study was approved by the IRB of AMC (IRB No. 2015-1368). The need for informed consent was waived by the ethics committee on the basis that this study utilized routinely collected medical data that were anonymously managed at all stages, including data cleaning and statistical analyses.

Results

Basic Characteristics

A total of 305 initial AE reports from 74 clinical trials were included in this study (see Table 1). Most initial AE reports

subjected to analysis occurred in sponsor-initiated trials (286/305, 93.8%) and multisite trials (273/305, 89.5%). Initial AE reports were more common in phase 3 trials, followed by phase 1 and 2 trials; approximately 70% of initial AE reports were from risk-level 3 trials (210/305, 68.9%). Most of the basic characteristics of the clinical trials were not significantly different between the pre- and postinitiative periods. The frequency of global trials was higher in the postinitiative period.

Table 1. Basic characteristics of initial adverse event (AE) reports from 74 clinical trials.

Category	Total (N=305), n (%)	Preinitiative (N=117), n (%)	Postinitiative (N=188), n (%)	*P* value[a]
Type of clinical trial				.46
Investigator-initiated trial	19 (6.2)	9 (7.6)	10 (5.3)	
Sponsor-initiated trial	286 (93.8)	109 (92.4)	177 (94.7)	
Number of sites involved				.06
Multisite	273 (89.5)	110 (93.2)	163 (87.2)	
Single site	32 (10.5)	7 (5.9)	25 (13.4)	
Phase				.08
1	78 (25.6)	24 (20.3)	54 (28.9)	
2	56 (18.4)	26 (22.0)	30 (16.0)	
3	117 (38.4)	33 (28.0)	84 (44.9)	
4 and other	54 (17.7)	34 (28.8)	20 (10.7)	
Scope of trial				.009
Domestic	48 (15.7)	27 (22.9)	21 (11.2)	
Global	257 (84.3)	90 (76.3)	167 (89.3)	
Risk level				.32
1	28 (9.2)	11 (9.3)	17 (9.1)	
2	67 (22.0)	20 (16.9)	47 (25.1)	
3	210 (68.9)	86 (72.9)	124 (66.3)	
Number of clinical trials	74 (24.3)	41 (34.7)	53 (28.3)	

[a]Chi-square test.

Adverse Event Awareness Performance

All three AE awareness performance indicators were significantly lower in the postinitiative phase (see Table 2 and Figure 2). However, due to the skewed distribution (see Figure 2), median values were 0 days in both the pre- and postinitiative phases. Statistical testing found the distributions to be significantly lower, and examination of the mean rank demonstrated that reporting times were lower and more consistent overall in the postinitiative phase. In particular, the onset to reporting showed a statistically significant difference (preinitiative: median 1 day, mean rank 173.82, vs postinitiative: median 1 day, mean rank 140.04, $P<.001$). The onset-to-awareness and awareness-to-reporting time also showed significant reductions between the pre- and postinitiative periods (preinitiative: median 0 days, mean rank 164.15, vs postinitiative: median 0 days, mean rank 146.06, $P=.04$; and preinitiative: median 1 day, mean rank 173.82, vs postinitiative: median 1 day, mean rank 140.04, $P=.02$, respectively).

Figure 2 shows the date difference between pre- and postinitiatives by three AE awareness indicators. For all three AE awareness indicators, the phenomenon of AE records taking more than 25 business days preinitiative disappeared postinitiative.

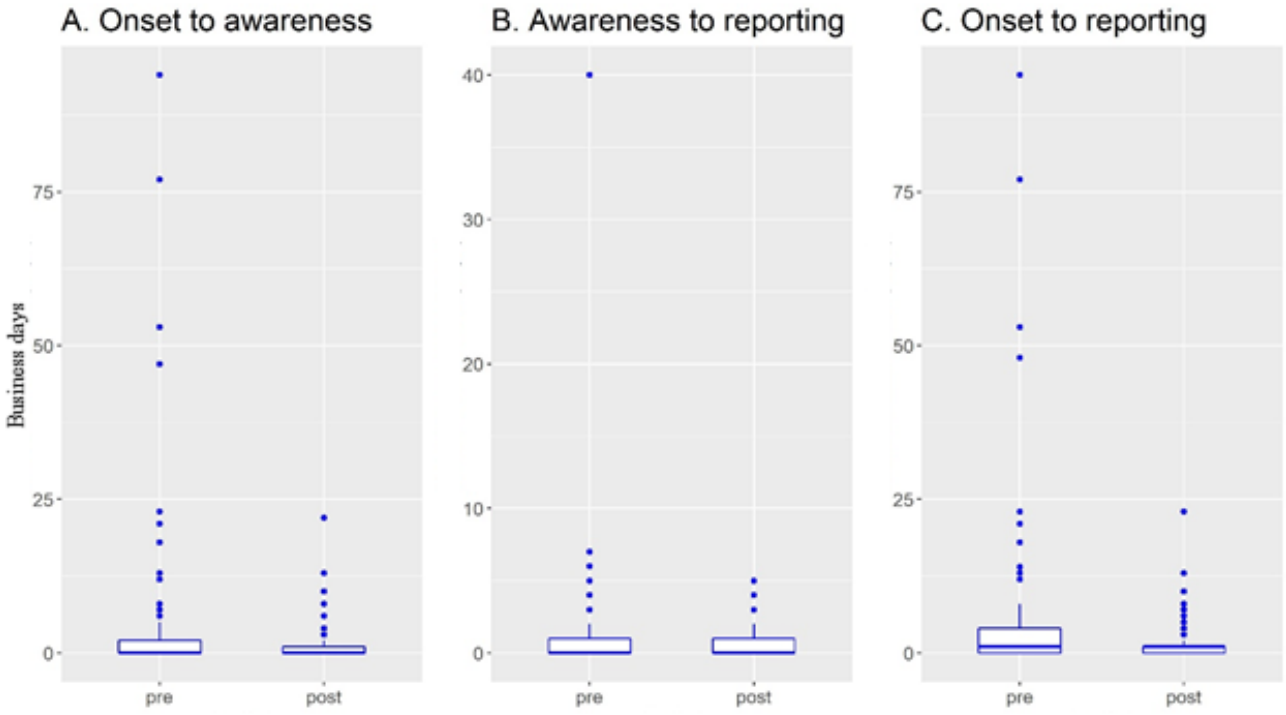
Table 2. Adverse event (AE) awareness efficiency, preinitiative and postinitiative.

AE awareness performance indicator	Preinitiative (N=118 AEs)		Postinitiative (N=188 AEs)		*P* value[a]	Effect size[b]
	Business days, median (IQR)	Mean rank (SD)	Business days, median (IQR)	Mean rank (SD)		
Onset to awareness	0 (0-2)	164.15 (83.41)	0 (0-1)	146.06 (68.67)	.04	0.117
Awareness to reporting	0 (0-1)	165.45 (79.58)	0 (0-1)	145.25 (67.36)	.02	0.135
Onset to reporting	1 (0-4)	173.82 (91.07)	1 (0-1)	140.04 (75.35)	<.001	0.197

[a]Wilcoxon rank-sum test.

[b]Effect sizes—0.1 (small effect), 0.3 (moderate effect), and 0.5 and above (large effect)—were calculated by dividing the absolute standardized test statistic *z* by the square root of the number of pairs.

Figure 2. Date difference comparison between preinitiative (pre) and postinitiative (post) periods by adverse event (AE) awareness performance indicators.



Adverse Event Awareness Performance by Phase and Risk

The phase subgroup analysis showed that all AE awareness indicators were reduced from the preinitiative period (see Table 3). The awareness-to-reporting and onset-to-reporting indicators in phase 1 studies showed statistically significant differences in median values and mean rank pre- and postinitiative (preinitiative: median 1 day, mean rank 47.94, vs postinitiative: median 0 days, mean rank 35.75, *P*=.01; and preinitiative: median 1 day, mean rank 47.4, vs postinitiative: median 1 day, mean rank 35.99, *P*=.03, respectively). In phase 4, there was no statistically significant difference in the onset-to-reporting timeline, but the SD of postinitiative ranks was significantly smaller than that of preinitiative ranks (preinitiative: rank SD 16.41 vs postinitiative: rank SD 11.89). Phase 3 and 4 studies showed greater differences than did phase 1 and 2 studies in the pre- and postinitiative SD values of all AE awareness performance indicators.

In the risk-level subgroup analysis, decreases of all AE report timelines were observed between pre- and postinitiative periods (see Table 4). Onset to awareness and onset to reporting of low-risk trials (level 1) showed significant reductions in the median and mean rank (preinitiative: median 4 days, mean rank 18.91, vs postinitiative: median 0 days, mean rank 11.65, *P*=.02; and preinitiative: median 4 days, mean rank 18.73, vs postinitiative: median 1 day, mean rank 11.76, *P*=.02, respectively). In high-risk trials (level 3), the distributions of awareness to reporting and onset to reporting were significantly different pre- and postinitiative (preinitiative: median 0 days, mean rank 115.94, vs postinitiative: median 0 days, mean rank 98.26, *P*=.02; and preinitiative: median 1 day, mean rank 117.36, vs postinitiative: median 1 day, mean rank 97.27, *P*=.01, respectively).

Table 3. Adverse event (AE) awareness efficiency, preinitiative and postinitiative, by study phase.

AE awareness performance indicator by phase	Preinitiative			Postinitiative			*P* value[a]	Effect size[b]
	Business days, median (IQR)	Mean rank (SD)	Number of records	Business days, median (IQR)	Mean rank (SD)	Number of records		
Phase 1 (n=22 studies, n=78 AEs)								
Onset to awareness	0 (0-1)	42.04 (19.86)	24	0 (0-0.75)	38.37 (17.04)	54	.41	0.09
Awareness to reporting	1 (0-1)	47.94 (20.61)		0 (0-1)	35.75 (18.55)		.01	0.28
Onset to reporting	1 (0.75-2.50)	47.4 (21.56)		1 (0-1)	35.99 (20.24)		.03	0.25
Phase 2 (n=14 studies, n=56 AEs)								
Onset to awareness	0 (0-1.75)	28.67 (16.42)	26	0.5 (0-1)	28.35 (13.65)	30	.94	0.01
Awareness to reporting	1 (0-1.75)	31.19 (15.85)		0 (0-1)	26.17 (13.78)		.21	0.17
Onset to reporting	1.5 (0.25-4.75)	31.98 (17.07)		1 (0-2)	25.48 (14.38)		.13	0.20
Phase 3 (n=29 studies, n=117 AEs)								
Onset to awareness	0 (0-1)	60.98 (31.18)	33	0 (0-1)	58.22 (26.28)	84	.63	0.04
Awareness to reporting	0 (0-1)	61.12 (26.84)		0 (0-0)	58.17 (24.54)		.57	0.05
Onset to reporting	0 (0-1)	61.71 (34.64)		0 (0-1)	57.93 (29.55)		.55	0.05
Phase 4 and other (n=9 studies, n=54 AEs)								
Onset to awareness	0.5 (0-3.75)	29.68 (15.20)	34	0 (0-1)	23.80 (12.03)	20	.15	0.20
Awareness to reporting	0 (0-0)	27.72 (12.41)		0 (0-0.25)	27.13 (10.88)		.87	0.02
Onset to reporting	1 (0-6.75)	29.93 (16.41)		1 (0-1)	23.38 (11.89)		.13	0.21

[a]Wilcoxon rank-sum test.

[b]Effect sizes—0.1 (small effect), 0.3 (moderate effect), and 0.5 and above (large effect)—were calculated by dividing the absolute standardized test statistic *z* by the square root of the number of pairs.

Table 4. Adverse event (AE) awareness efficiency in preinitiative and postinitiative periods according to risk level.

AE awareness performance indicator by risk level	Preinitiative			Postinitiative			*P* value[a]	Effect size[b]
	Business days, median (IQR)	Mean rank (SD)	Number of records	Business days, median (IQR)	Mean rank (SD)	Number of records		
Level 1 (n=5 studies, n=28 AEs)								
Onset to awareness	4 (0.5-22.0)	18.91 (8.09)	11	0 (0-1)	11.65 (5.99)	17	.02	0.46
Awareness to reporting	0 (0-0)	14.23 (6.10)		0 (0-0)	14.68 (5.90)		.87	0.03
Onset to reporting	4 (1-23)	18.73 (8.35)		1 (0-1)	11.76 (6.30)		.02	0.43
Level 2 (n=17 studies, n=67 AEs)								
Onset to awareness	0.5 (0-6.25)	39.75 (20.31)	20	0 (0-1)	31.55 (15.27)	47	.08	0.22
Awareness to reporting	0 (0-0)	33.73 (12.80)		0 (0-0)	34.12 (12.54)		.92	0.01
Onset to reporting	0.5 (0-6.25)	37.65 (21.93)		1 (0-1)	32.45 (16.00)		.28	0.13
Level 3 (n=52 studies, n=210 AEs)								
Onset to awareness	0 (0-1)	108.20 (53.50)	86	0 (0-1)	103.63 (47.69)	124	.52	0.05
Awareness to reporting	0 (0-1)	115.94 (56.72)		0 (0-1)	98.26 (48.93)		.02	0.16
Onset to reporting	1 (0-3)	117.36 (61.19)		1 (0-1)	97.27 (53.62)		.01	0.17

[a]Wilcoxon rank-sum test.

[b]Effect sizes—0.1 (small effect), 0.3 (moderate effect), and 0.5 and above (large effect)—were calculated by dividing the absolute standardized test statistic *z* by the square root of the number of pairs.

Discussion

Principal Findings

The most notable finding of this study is that the timeline from patients' unscheduled visits (ie, onset), due to AE, to safety reporting was significantly reduced after the implementation of the AEAS. Also, the variability in the amount of time between patient visits and investigator awareness was lower after the implementation. This suggests that the AEAS notifications are effective for improving the speed with which investigators or clinical research coordinators are informed, thereby allowing them to take prompt action against AE occurrences.

There have been several approaches to enhance the efficiency, completeness, and consistency of safety reporting through the use of Web-based electronic safety reporting modules [18,19]. However, such approaches do not detect patients' unscheduled visits at the site level. To our knowledge, this is the first study to evaluate the effectiveness of the implementation of an IT-driven AE awareness system for clinical trials at the site level. Early safety signal awareness with the alert system may contribute to better patient protection after the occurrence of AEs. A recent study found that many SAE results registered in ClinicalTrials.gov were yet to be published or omitted from publications [20]. Due to the imbalance of information on AE reporting, there is a high demand for more comprehensive approaches to ensure the safety of clinical trial participants [20-22]. There are only a few reports that have evaluated the timeline between AE onset and the initial reporting of the AE at the site level. One study reported that the mean duration from onset to reporting was 25 days for nonserious AEs and 11 days for SAEs [18]. However, to the best of our knowledge, there are no reports on other systems designed to improve the detection of AEs during clinical trials. Caution is necessary when comparing our findings with those of other studies because of the differences in protocols, safety reporting systems, and sites. However, our improved safety reporting metrics—from unscheduled visit to reporting—which were less than 2 days in the postinitiative period, deserve highlighting.

Another interesting finding is that the improvement in turnaround times after AEAS implementation was more prominent in phase 1 and in low- and high-risk clinical trials. The primary goal of a phase 1 study is to identify safety profiles and determine the dose-limiting toxicities of a new drug. This might be due to the increased attention being paid to participants in phase 1; therefore, the improvement in the reporting timeline was not as noticeable in trials of other phases as it was in phase 1 trials. We note that the safety reporting timeline is shorter postinitiative than preinitiative in all phases of clinical research.

Limitations

The main limitation of our study is that our study results are based on a retrospective analysis of clinical trials carried out in a single academic medical center. Also, this system could not detect patient visits to other hospitals, which would require patient self-reporting; the rate of patient self-reporting may be improved with education, telephone monitoring, and questioning during trial visits. Another limitation is that we only analyzed the timeline for reporting AEs. Nevertheless, detailed analyses of AEs in clinical trials were not accessible at this point, as such analysis is only possible when the data are reported to the regulatory agency and published. Due to the study's retrospective nature, there is a trend of imbalances among several characteristics, such as phase of trial or number of sites.

Conclusions

In this study, we demonstrated that the AEAS, a CTMS-driven real-time automatic alert system, can effectively improve safety reporting timelines. The AEAS resulted in overall reductions in AE awareness timelines in all clinical trials, especially in phase 1 trials and low- and high-risk studies. These findings suggest that IT-driven automatic alert systems are effective in improving safety reporting timelines, which may ultimately enhance patient safety.

Acknowledgments

This study was supported by a grant from the Korean Health Technology R&D Project through the Korea Health Industry Development Institute funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI18C2383, HI14C1061). We thank Dr Joon Seo Lim from the Scientific Publications Team at the AMC.

Authors' Contributions

YRP, HYK, SP, and TWK conceived and designed the study. YRP and SP reviewed records and collected the data. YRP, SB, and HRK analyzed the data. YRP, HYK, and YKY wrote the manuscript. YSL and TWK reviewed the manuscript.

Conflicts of Interest

None declared.

References

1. Sheiner LB. Learning versus confirming in clinical drug development. Clin Pharmacol Ther 1997 Mar;61(3):275-291. [doi: 10.1016/S0009-9236(97)90160-0] [Medline: 9084453]
2. Owens P, Raddad E, Miller JW, Stille JR, Olovich KG, Smith NV, et al. A decade of innovation in pharmaceutical R&D: The Chorus model. Nat Rev Drug Discov 2015 Jan;14(1):17-28. [doi: 10.1038/nrd4497] [Medline: 25503514]
3. Berlin JA, Glasser SC, Ellenberg SS. Adverse event detection in drug development: Recommendations and obligations beyond phase 3. Am J Public Health 2008 Aug;98(8):1366-1371. [doi: 10.2105/AJPH.2007.124537] [Medline: 18556607]
4. Wood AJ, Stein CM, Woosley R. Making medicines safer--The need for an independent drug safety board. N Engl J Med 1998 Dec 17;339(25):1851-1854. [doi: 10.1056/NEJM199812173392512] [Medline: 9854125]
5. Curfman GD, Morrissey S, Drazen JM. Blueprint for a stronger Food and Drug Administration. N Engl J Med 2006 Oct 26;355(17):1821. [doi: 10.1056/NEJMe068237] [Medline: 17030844]
6. FDA Adverse Event Reporting System (FAERS): Latest Quarterly Data Files. Silver Spring, MD: US Food and Drug Administration; 2018. URL: https://www.fda.gov/drugs/fda-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-latest-quarterly-data-files [accessed 2019-09-03]
7. Diao G, Liu GF, Zeng D, Wang W, Tan X, Heyse JF, et al. Efficient methods for signal detection from correlated adverse events in clinical trials. Biometrics 2019 Sep;75(3):1000-1008. [doi: 10.1111/biom.13031] [Medline: 30690717]
8. Ioannidis JP, Evans SJ, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, CONSORT Group. Better reporting of harms in randomized trials: An extension of the CONSORT statement. Ann Intern Med 2004 Nov 16;141(10):781-788. [doi: 10.7326/0003-4819-141-10-200411160-00009] [Medline: 15545678]
9. Sivendran S, Latif A, McBride RB, Stensland KD, Wisnivesky J, Haines L, et al. Adverse event reporting in cancer clinical trial publications. J Clin Oncol 2014 Jan 10;32(2):83-89. [doi: 10.1200/JCO.2013.52.2219] [Medline: 24323037]
10. Park YR, Yoon YJ, Koo H, Yoo S, Choi C, Beck S, et al. Utilization of a clinical trial management system for the whole clinical trial process as an integrated database: System development. J Med Internet Res 2018 Apr 24;20(4):e103 [FREE Full text] [doi: 10.2196/jmir.9312] [Medline: 29691212]
11. Ryu HJ, Kim WS, Lee JH, Min SW, Kim SJ, Lee YS, et al. Asan medical information system for healthcare quality improvement. Healthc Inform Res 2010 Sep;16(3):191-197 [FREE Full text] [doi: 10.4258/hir.2010.16.3.191] [Medline: 21818439]
12. Ahn HS, Kim HJ. Development and implementation of clinical practice guidelines: Current status in Korea. J Korean Med Sci 2012 May;27 Suppl:S55-S60 [FREE Full text] [doi: 10.3346/jkms.2012.27.S.S55] [Medline: 22661872]
13. Englev E, Petersen KP. ICH-GCP Guideline: Quality assurance of clinical trials. Status and perspectives [Article in Danish]. Ugeskr Laeger 2003 Apr 14;165(16):1659-1662. [Medline: 12756823]
14. Wagner MM, Moore AW, Aryel RM, editors. Handbook of Biosurveillance. London, UK: Elsevier Academic Press; 2006.
15. "What to Expect of a Regulatory Inspection": Informational Handout for Farmers. College Park, MD: US Food and Drug Administration, Center for Food Safety and Applied Nutrition URL: https://www.fda.gov/media/124328/download [accessed 2020-01-12]
16. Guidance for Industry: Postmarketing Adverse Event Reporting for Nonprescription Human Drug Products Marketed Without an Approved Application. Silver Spring, MD: US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER); 2009 Jul. URL: https://www.fda.gov/media/77193/download [accessed 2020-01-09]

17. Brosteanu O, Houben P, Ihrig K, Ohmann C, Paulus U, Pfistner B, et al. Risk analysis and risk adapted on-site monitoring in noncommercial clinical trials. Clin Trials 2009 Dec;6(6):585-596. [doi: 10.1177/1740774509347398] [Medline: 19897532]
18. Zhao W, Waldman BD, Dillon C, Pauls K, Kim J, Patterson L, et al. A Web-based medical safety reporting system for a large multicenter clinical trial: The ALIAS experience. Contemp Clin Trials 2010 Nov;31(6):536-543 [FREE Full text] [doi: 10.1016/j.cct.2010.08.010] [Medline: 20828636]
19. Perez RP, Finnigan S, Patel K, Whitney S, Forrest A. Clinical trial electronic portals for expedited safety reporting: Recommendations from the Clinical Trials Transformation Initiative Investigational New Drug Safety Advancement Project. JMIR Cancer 2016 Dec 15;2(2):e16 [FREE Full text] [doi: 10.2196/cancer.6701] [Medline: 28410179]
20. Tang E, Ravaud P, Riveros C, Perrodeau E, Dechartres A. Comparison of serious adverse events posted at ClinicalTrials.gov and published in corresponding journal articles. BMC Med 2015 Aug 14;13:189 [FREE Full text] [doi: 10.1186/s12916-015-0430-4] [Medline: 26269118]
21. Seruga B, Templeton AJ, Badillo FE, Ocana A, Amir E, Tannock IF. Under-reporting of harm in clinical trials. Lancet Oncol 2016 May;17(5):e209-e219. [doi: 10.1016/S1470-2045(16)00152-2] [Medline: 27301048]
22. Sivendran S, Latif A, McBride RB, Stensland KD, Wisnivesky J, Haines L, et al. Adverse event reporting in cancer clinical trial publications. J Clin Oncol 2014 Jan 10;32(2):83-89. [doi: 10.1200/JCO.2013.52.2219] [Medline: 24323037]

Abbreviations

ADAMON: ADApted MONitoring
AE: adverse event
AEAS: Adverse Event Awareness System
AMC: Asan Medical Center
CONSORT: Consolidated Standards of Reporting Trials
CTMS: clinical trial management system
EMR: electronic medical record
IRB: institutional review board
IT: information technology
SAE: serious adverse event
SUSAR: suspected unexpected serious reaction

Edited by G Eysenbach; submitted 13.04.19; peer-reviewed by J Rössler, P Wicks; comments to author 14.06.19; revised version received 09.09.19; accepted 17.12.19; published 27.02.20.

Please cite as:
Park YR, Koo H, Yoon YK, Park S, Lim YS, Baek S, Kim HR, Kim TW
Expedited Safety Reporting Through an Alert System for Clinical Trial Management at an Academic Medical Center: Retrospective Design Study
JMIR Med Inform 2020;8(2):e14379
URL: http://medinform.jmir.org/2020/2/e14379/
doi:10.2196/14379
PMID:

Original Paper

Explanatory Model of Dry Eye Disease Using Health and Nutrition Examinations: Machine Learning and Network-Based Factor Analysis From a National Survey

Sang Min Nam[1*], MD, PhD; Thomas A Peterson[2*], PhD; Atul J Butte[2], MD, PhD; Kyoung Yul Seo[3], MD, PhD; Hyun Wook Han[4], MD, PhD

[1]Department of Ophthalmology, CHA Bundang Medical Center, CHA University, Seongnam, Republic of Korea

[2]Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, CA, United States

[3]Department of Ophthalmology, Institute of Vision Research, Eye and Ear Hospital, Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea

[4]Department of Biomedical Informatics, CHA University School of Medicine, CHA University, Seongnam, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Hyun Wook Han, MD, PhD
Department of Biomedical Informatics, CHA University School of Medicine, CHA University
335 Pangyo-ro
Seongnam, 13488
Republic of Korea
Phone: 82 318817109
Email: hwhan@chamc.co.kr

Abstract

Background: Dry eye disease (DED) is a complex disease of the ocular surface, and its associated factors are important for understanding and effectively treating DED.

Objective: This study aimed to provide an integrative and personalized model of DED by making an explanatory model of DED using as many factors as possible from the Korea National Health and Nutrition Examination Survey (KNHANES) data.

Methods: Using KNHANES data for 2012 (4391 sample cases), a point-based scoring system was created for ranking factors associated with DED and assessing patient-specific DED risk. First, decision trees and lasso were used to classify continuous factors and to select important factors, respectively. Next, a survey-weighted multiple logistic regression was trained using these factors, and points were assigned using the regression coefficients. Finally, network graphs of partial correlations between factors were utilized to study the interrelatedness of DED-associated factors.

Results: The point-based model achieved an area under the curve of 0.70 (95% CI 0.61-0.78), and 13 of 78 factors considered were chosen. Important factors included sex (+9 points for women), corneal refractive surgery (+9 points), current depression (+7 points), cataract surgery (+7 points), stress (+6 points), age (54-66 years; +4 points), rhinitis (+4 points), lipid-lowering medication (+4 points), and intake of omega-3 (0.43%-0.65% kcal/day; −4 points). Among these, the age group 54 to 66 years had high centrality in the network, whereas omega-3 had low centrality.

Conclusions: Integrative understanding of DED was possible using the machine learning–based model and network-based factor analysis. This method for finding important risk factors and identifying patient-specific risk could be applied to other multifactorial diseases.

(JMIR Med Inform 2020;8(2):e16153) doi:10.2196/16153

KEYWORDS

dry eye disease; epidemiology; machine learning; systems analysis; patient-specific modeling

Introduction

Background and Related Studies

Dry eye disease (DED) is defined as "a multifactorial disease of the ocular surface characterized by a loss of homeostasis of the tear film, and accompanied by ocular symptoms" [1]. Due to its multifactorial etiology, DED cannot be characterized by a single process and its management is complicated, in which finding the major causative factors behind DED is critical to appropriate treatment [1]. Therefore, identification of DED-related factors may enable advances in diagnosis, elucidative pathophysiology, therapy, and public education, as well as improvement of general and ocular health [2]. Indeed, various nonmodifiable, modifiable, environmental, and medical factors related to DED have been reported by observational studies and population-based, cross-sectional epidemiological studies [2]. DED risk factors are categorized as consistent, probable, and inconclusive; age, sex, Meibomian gland dysfunction (MGD), connective tissue disease, Sjogren syndrome, androgen deficiency, computer use, contact lens wear, estrogen replacement therapy, and medication use (eg, antihistamines, antidepressants, and anxiolytics) are identified as consistent risk factors [2].

Previously, a limited number of DED-associated factors were investigated using the Korea National Health and Nutrition Examination Survey (KNHANES) [3]. Although KNHANES consists of a large number of variables from health interview questionnaires, health examinations, and nutrition surveys, they were not fully utilized [3]. In addition, previous studies on DED have identified DED-related factors, instead of building a DED model to assess the risk of DED for new individuals [3-7].

Highlights of This Study

In this study, we generated a point-based model with DED-associated factors from KNHANES using machine learning algorithms and Lasso regularization. These methods can improve the model performance to predict DED by selecting features from a large number of variables from a large dataset without overfitting while preserving complex interactions among features [8]. Furthermore, interactions among the factors were explored by network analysis. When the network analysis was applied to the model, a systemic understanding of DED, which cannot be achieved by conventional methods, was possible by showing the linkages between the relevant factors. To the best of our knowledge, this was the first attempt at building a machine learning–based model to evaluate the individual risks of DED and visualize the state using the network graph of DED-associated factors.

Methods

Overview of Survey Data

The design, methods, and data resource profile of KNHANES are available on the Web and in publications [9-11]. In short, KNHANES is an annual survey performed by the Korea Centers for Disease Control and Prevention (KCDC) in the Republic of Korea, which assesses the health and nutritional status of the population [10]. KNHANES is a nationwide cross-sectional survey of a representative set of 10,000 noninstitutionalized civilian individuals who are aged 1 year and older. Both DED assessment and food frequency surveys were conducted only in 2012. In the 2012 KNHANES, 192 primary sampling units (PSUs) were drawn from about 200,000 geographically defined PSUs nationwide; 20 final target households were sampled for each PSU as secondary sampling units [9]. KNHANES V (2012) was approved by the KCDC Research Ethics Committee (2012-01EXP-01-2C), and written informed consent was obtained from all subjects.

Variable Inclusion

Four data files, HN12_ALL (health examination, health survey, and nutrient survey), HN12_ENT (ear, nose, and throat examination), HN12_EYE (eye examination), and HN12_FFQ (food frequency survey), were combined. DED was considered to be present when a subject had been diagnosed with DED by an ophthalmologist (the variable *E_DES_dg*) and was experiencing dryness (*E_DES_ds*). Conversely, patients were defined as DED-negative in the absence of both a diagnosis and symptoms. *E_DES_dg* and *E_DES_ds* are available for persons who are aged 19 years and older [11].

The included variables are listed in Textbox 1, and the overall analysis is summarized in Figure 1.

All variables were available for subjects aged 19 years and older except those of food frequency (19-64 years) and osteoarthritis radiology (≥50 years) [9]. The LDL level was calculated using the Friedewald equation, LDL=total cholesterol−(HDL+TG/5), with exclusion of TG levels of higher than 400 mg/dL [12].

Textbox 1. Included study variables of the Korea National Health and Nutrition Examination Survey data (2012).

Health examination data

Physical examination

- Body mass index (BMI), rhinitis, sinusitis, blepharoptosis [11], and cataract

Blood test results

- Anemia, hemoglobin, hematocrit, iron, total iron-binding capacity, ferritin, hemoglobin A_{1c}, white blood cell count, platelet count, red blood cell count, aspartate aminotransferase, alanine aminotransferase, creatinine, urea nitrogen, and vitamin D

Fasting (≥8 hours) blood parameters

- Sugar level, low-density lipoprotein cholesterol (LDL) level, high-density lipoprotein cholesterol (HDL) level, triglyceride (TG) level

Hypercholesterolemia definition: total cholesterol (TC)≥240 mg/dL or lipid-lowering medication

Hypertension definition: systolic blood pressure≥140 mm Hg, or diastolic blood pressure ≥ 90 mm Hg, or medication

Diabetes mellitus definition: fasting blood sugar level≥126 mg/dL, or diagnosis, or medication, or insulin injection

Fundus photography

- age-related macular degeneration, diabetic retinopathy

Osteoarthritis on radiology

Health survey data

- Age, educational stage, occupational class, household income, weight changes in the last year, mean duration of sleep per day, stress recognition, current smoker, frequency of drinking alcohol, activity level, lipid-lowering medications, diagnosed glaucoma, eye surgery, and menstruation

Diagnosed current diseases

- dyslipidemia, depression, stroke, myocardial infarction or angina, rheumatoid arthritis, thyroid disease, atopic dermatitis, and asthma

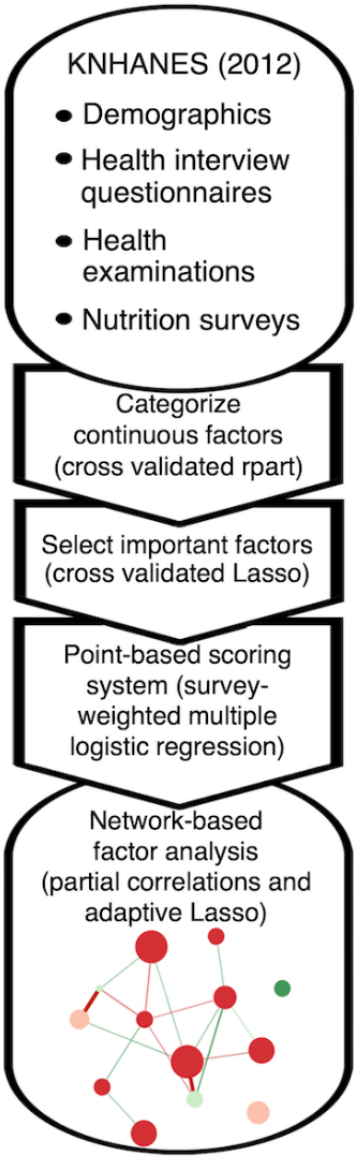
Diagnosed cancers

- stomach, colon, breast, cervix, and thyroid

Food frequency survey data (daily intake)

- Energy, carbohydrate, protein, total fat, n-3 polyunsaturated fatty acid, n-6 polyunsaturated fatty acid, saturated fatty acid, cholesterol, fiber, vitamin A, vitamin B1, vitamin B2, vitamin C, niacin, iron, calcium, potassium, phosphorus, and sodium

Figure 1. Flowchart of overall analysis steps. KNHANES: Korea National Health and Nutrition Examination Survey.



Subsampling of Training and Test Sets and Categorization of Factors

Most DED cases (80.00%, 3513/4391) were used as training, and the other cases were used for testing. Likewise, non-DED cases were subsampled into training and test sets in the same way. Next, the categorization or recategorization of the factors was performed for the training set in consideration of reference values. Here, optimal cutoffs were determined by training a decision tree on the training data and using binarized decision tree rules as factors in the final regression model [13,14]. Missing values of each variable were classified as a separate class.

Factor Selection Using Lasso (Least Absolute Shrinkage and Selection Operator)

Factors were transformed into dummy-coded variables, in which the largest category was used as reference and was excluded during model construction, and missing values were not included in the Lasso procedure.

Lasso trained using cross validation was applied to the transformed dummy variables with area under the curve (AUC) as a stopping metric and *wt_tot* as the sample weight for the analysis of the associations between the health interview, health examination, and nutrition survey. To regularize the model, we selected the optimal lambda using cross validation (*lambda.1se* in glmnet, ie, the lambda that yields an error one standard error away from the minimum error).

Construction of a Model for Dry Eye Disease

Using the lasso-selected factors, a survey-weighted multiple logistic regression model was constructed from the complex survey design of KNHANES. The survey design was represented using the variable *psu* for PSU and *ID_fam* for the secondary sampling unit, *kstrata* for strata, and *wt_tot* for weights.

Developing a Point-Based Scoring System for Dry Eye Disease

A point-based scoring system was developed by multiplying the coefficients of factors in the survey-weighted regression model by 10 and rounding to the nearest integer [15]. The total

score of each individual in the training set was determined by summing the points for factors accurately describing that individual. Next, performance was assessed using weighted receiver-operating characteristic (ROC) curves and the AUCs with survey sample weight (*wt_tot*). An optimal cutoff for the point-based system was determined by maximizing Youden's index value (sensitivity+specificity−1).

Testing the Point-Based Scoring System for Dry Eye Disease

The model's performance was assessed using the test set. The AUC's confidence interval was calculated; sensitivity and specificity were reported using the point-based system's cutoff determined from the training set.

Analysis of Dry Eye Disease-Risk Factors

A survey-weighted multiple logistic regression analysis was performed using the factors selected by lasso. Odds ratios (ORs) were calculated by exponentiating the coefficient derived by logistic regression. Estimated population counts and proportions for categories were computed.

Network Analysis of Dry Eye Disease-Associated Factors

With the training set, a correlation matrix for the DED-associated factors was created. Weighted Pearson correlation coefficients between two variables were calculated. Next, a network graph was plotted by setting the graph argument to "glasso" and the layout to "spring." A partial correlation network was drawn using the graphical lasso algorithm and the Extended Bayesian Information Criterion by which false positive edges were controlled. Each edge represents the relationship between 2 nodes after controlling for all other relationships in the network [16,17]. The Fruchterman-Reingold algorithm is applied with the "spring" layout, in which the lengths of edges are dependent on their absolute weights [16]. Green edges indicate positive weights (correlations) and red edges indicate negative weights. Color saturation and edge width correspond to the absolute weight relative to the strongest weight in the graph. Node size was proportional to the *z*-score for the absolute point of the factor. Nodes were grouped as *significant* ($P<.05$, risk factor analysis) or *possible* ($P\geq.05$, risk factor analysis).

Three centrality indices (strength, closeness, and betweenness) were computed. Centrality is the absolute sum of the edge weights connected to the node, closeness is the sum of the shortest distances from the node to all other nodes in the network, and betweenness is the number of times in which the node lies on the shortest path between 2 other nodes [17,18].

Statistics and Software

R version 3.6.1 and variable functions from its packages were used: decision tree, "rpart" in the caret package (using down-sampling and cross-validation) [19,20]; dummy-coded variables, "dummy.code" in the psych package [21]; cross-validation for Lasso, "cv.glmnet" in the glmnet package [22]; survey-weighted multiple logistic regression, "svyglm" in the survey package [23]; weighted ROC curve and AUC, "WeightedROC" and "WeightedAUC," respectively, in the WeightedROC package [24]; confidence interval of AUC, "withReplicates" in the survey package [23]; estimation of population counts and proportions, the survey package [23]; general graphs, the ggplot2 package [25]; weighted correlation, "wtd.cor" in the weights package [26]; network graph, "qgraph" in the qgraph package [16]; and centrality indices, "qgraph" and "centralityTable" in the qgraph package [16].

Results

Point-Based Scoring Model for Dry Eye Disease

Total sample sizes for DED and non-DED were 575 and 3816 cases, respectively. The estimated prevalence of DED was 10.5% (SE 1.0%): 5.3% (SE 1.0%) for men and 15.9% for women (SE 1.0%). A total of 13 factors were selected by lasso and the point-based scoring system for each factor is outlined in Table 1.

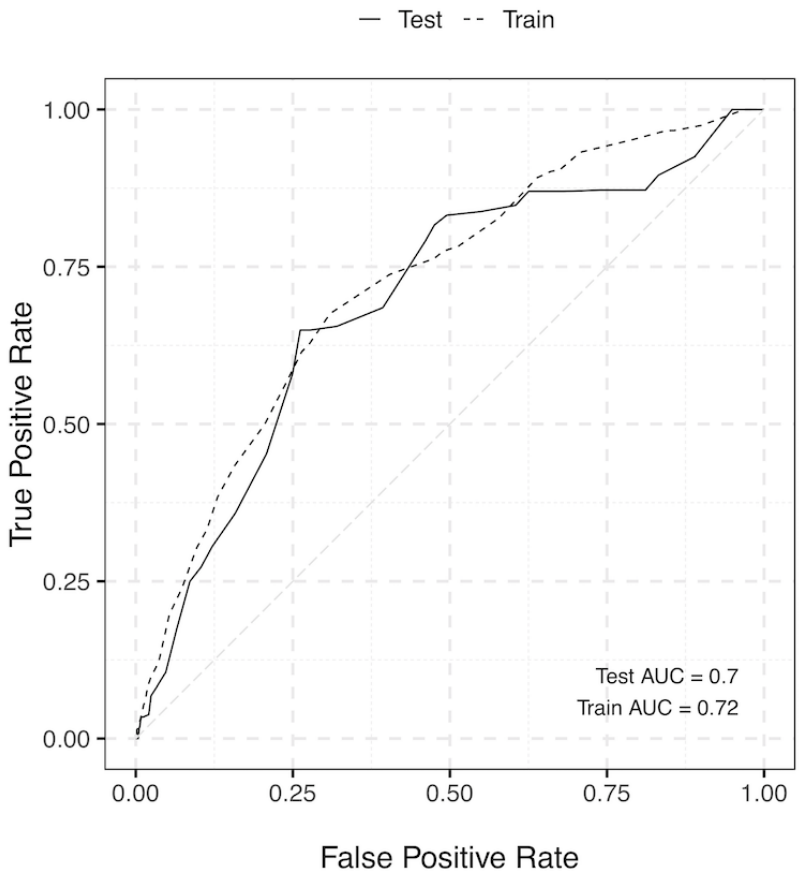
Using this scoring system on the test set achieved an AUC of 0.70 (95% CI 0.61-0.78; Figure 2). Sensitivity and the specificity were 0.66 and 0.68, respectively, at a cutoff of 10 points.

Table 1. Point-based scoring system for assessing individual risk of dry eye disease using coefficients from a survey-weighted multiple logistic regression model.

Variables	Regression coefficient (beta)	Standard error	Points[a]
Women	.865	0.182	9
Corneal refractive surgery	.903	0.281	9
Current depression	.709	0.294	7
Eye surgery: cataract	.705	0.196	7
Perceived stress: much to extreme	.560	0.136	6
Other ocular surgeries	.646	0.258	6
Phosphorus intake <746 mg/day	.454	0.182	5
Age 54-66 years	.396	0.179	4
Rhinitis by physical examination	.384	0.144	4
Lipid-lowering medications	.408	0.194	4
Cholesterol intake ≥240 mg/day	−.145	0.151	−1
Current smoker	−.441	0.242	−4
Omega-3 intake, 0.43%-0.65% kcal/day	−.407	0.172	−4

[a]Calculated by multiplying the coefficient of the variable by 10 and rounding to the nearest integer. A positive point means a positive predictor for dry eye disease. Dry eye disease is indicated when the sum of all points is 10 or higher.

Figure 2. Weighted receiver-operating characteristic (ROC) of the point-based scoring system for predicting dry eye disease. ROCs for train and test sets were compared. AUC: area under the curve.



Risk Factor Analysis For Dry Eye Disease

In the risk factor analysis, 10 of the 13 variables were significant (P<.05; Table 2). The top 3 significant risk factors in the point-based model were women, corneal refractive surgery, and current depression (Tables 1 and 2). Omega-3 intake between 0.43% (1003 mg for total 2100 kcal) and 0.65% (1517 mg for total 2100 kcal) was a significant protective factor.

Population counts (n), proportions (%), and ORs were estimated according to complex survey design. ORs and *P* values were calculated by multiple logistic regression including all listed variables. The missing data category for each variable were included for calculation but not shown in the table.

Table 2. Population counts (n), proportions (%), and odds ratios of variables in the points-based scoring system for dry eye disease.

Factors	Healthy, n (%)	Dry eye disease, n (%)	OR[a] (95% CI)	*P* value
Sex/gender				
Men	16,277,579 (54.19)	911,587 (25.90)	Reference	Reference
Women	13,761,300 (45.81)	2,607,754 (74.10)	2.6 (1.8-3.6)	<.001
Perceived stress				
None to a little	21,872,165 (72.81)	2,177,913 (61.88)	Reference	Reference
Much to extreme	6,864,096 (22.85)	1,202,940 (34.18)	1.6 (1.2-2.0)	<.001
Eye surgery				
None	26,673,187 (88.80)	2,701,904 (76.77)	Reference	Reference
Cataract	1,185,785 (3.95)	276,588 (7.86)	1.8 (1.2-2.6)	.004
Corneal refractive	1,150,662 (3.83)	339,857 (9.66)	2.8 (1.7-4.6)	<.001
Other	1,029,243 (3.43)	200,991 (5.71)	1.6 (1.0-2.5)	.08
Rhinitis on inspection				
No	21,295,695 (70.89)	2,180,596 (61.96)	Reference	Reference
Yes	8,047,298 (26.79)	1,243,634 (35.34)	1.5 (1.2-2.0)	.001
Lipid-lowering medications				
No	27,360,794 (91.08)	3,087,286 (87.72)	Reference	Reference
Yes	1,315,445 (4.38)	282,072 (8.01)	1.5 (1.1- 2.0)	.02
Age (year)				
<54 or ≥66	25,121,235 (83.63)	2,757,338 (78.35)	Reference	Reference
54-66	4,917,644 (16.37)	762,003 (21.65)	1.4 (1.1-2.0)	.02
Current depression				
No	25,591,681 (85.20)	2,616,142 (74.34)	Reference	Reference
Yes	456,573 (1.52)	153,162 (4.35)	1.9 (1.1-3.3)	.02
Omega-3 intake (kcal/day)				
<0.43% or >0.65%	13,804,921 (45.96)	1,812,313 (51.50)	Reference	Reference
0.43%-0.65%	10,414,321 (34.67)	857,985 (24.38)	0.7 (0.5-1.0)	.04
Phosphorus intake (mg/day)				
≥746	19,948,841 (66.41)	1,940,991 (55.15)	Reference	Reference
<746	4,270,401 (14.22)	729,307 (20.72)	1.4 (1.0-2.0)	.09
Current smoker				
No	21,334,370 (71.02)	2,950,840 (83.85)	Reference	Reference
Yes	7,399,648 (24.63)	430,013 (12.22)	0.7 (0.5-1.1)	.14
Cholesterol intake (mg/day)				
<240	12,037,627 (40.07)	1,575,001 (44.75)	Reference	Reference
≥240	12,181,614 (40.55)	1,095,296 (31.12)	0.9 (0.7-1.2)	.57

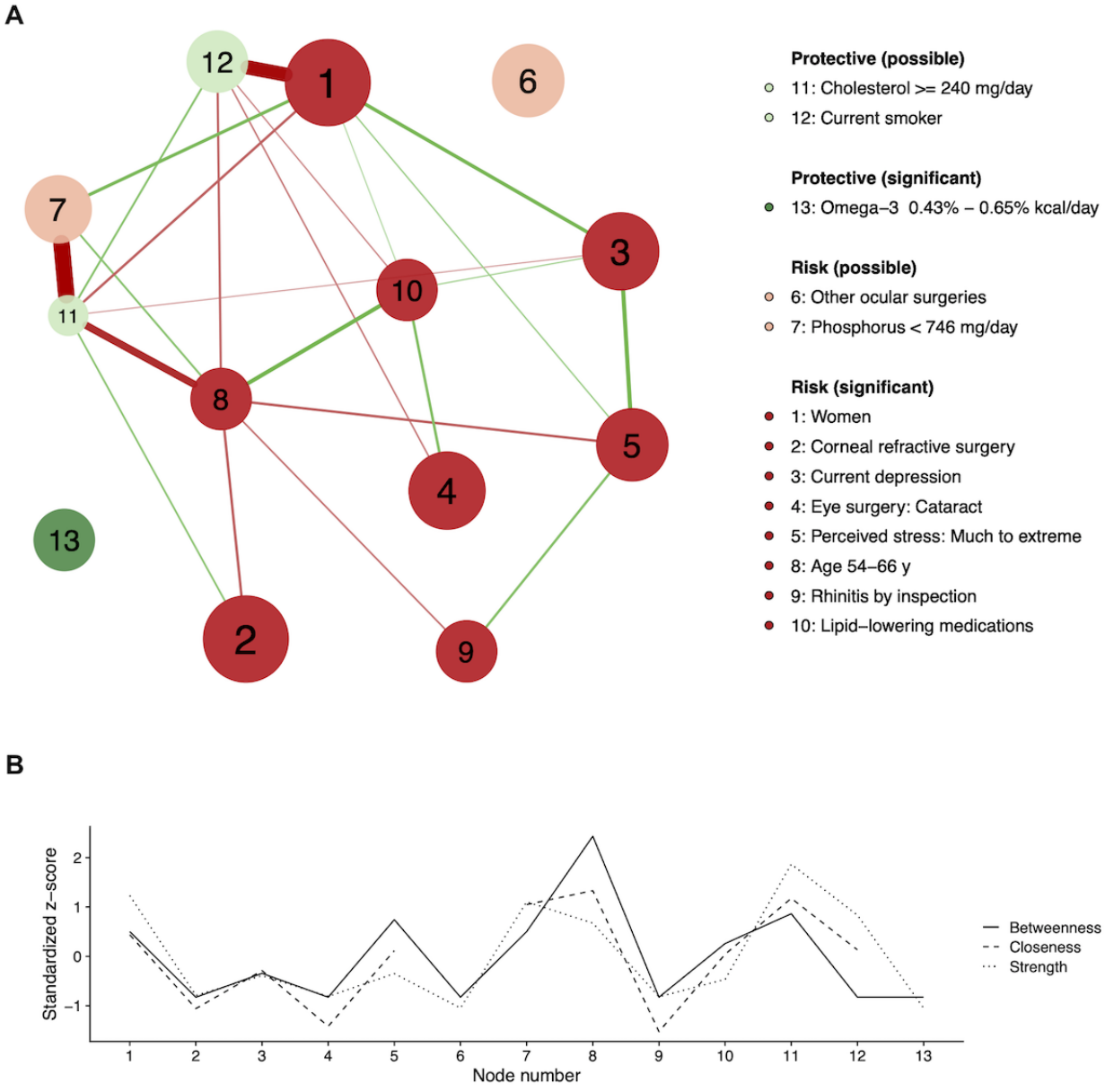
[a]OR: odds ratio.

Network Analysis for Dry Eye Disease Model

In Figure 3, model factors are depicted in a partial correlation network with centrality indices. The network-based factor analysis in Figure 3 allows for the interrogation of the interrelatedness of factors associated with DED, with larger nodes representing factors' importance (points), green nodes representing protective factors, and red nodes representing risk

factors. According to centrality indices (Figure 3), *Age 54-66y* (node 8) had high centrality in the network. For *other ocular surgeries* (node 6) and *omega-3* (node 13), the closeness indices were too low to calculate owing to lack of the connections to other nodes.

Figure 3. The partial correlation (adaptive LASSO) network (A) and the centrality indices (betweenness, closeness, and strength) (B) of the factors associated with the dry eye disease. Factors were positively (risk) or negatively (protective) associated with dry eye disease. Factors can be significant (P<.05) or possible (P≥.05) according to the risk factor analysis. The node size is proportional to the absolute value of the point for the node's variable. Green and red edges mean positive and negative connections, respectively. The edge with the highest absolute weight will have full color saturation and be the widest.



Four significant risk factors were linked in succession from *women* (node 1) to *current depression* (node 3), *much to extreme stress* (node 5), and *rhinitis* (node 9). Other serial connections were found in three significant risk factors, *age 54-66y*, *lipid-lowering medication* (node 10), and *cataract surgery* (node 4). Nonsignificant factors were strongly connected with other significant factors, for example, *current smoker* (node 12) to *women*, and *cholesterol intake* (node 11) to *age 54-66y*. Another nonsignificant factor, *phosphorous intake* (node 7), was closely associated with the *cholesterol intake*.

Discussion

Principal Findings

Our model showed moderate performance for DED prediction with a point-based scoring system in which the maximum AUC might reach 0.78. Our study chose a stricter definition of DED because the individuals were not only required to be symptomatic but also have a physician diagnosis. In addition, the absence of DED was rigorously defined as a lack of symptoms and no physician diagnosis in the past. According to the TFOS DEWS II report, DED is diagnosed on the basis of

the presence of a symptom and positivity for one or more homeostatic markers [1]. Our DED definition more reflected a diagnosis of DED, and thus, the prevalence could be lower than that of prior studies that used a symptomatic definition [2]. However, even our definition was imperfect because diagnostic tests were not performed and might be biased by the availability of a clinic in the local area or by the respondent's condition. This may explain, in part, the moderate diagnostic performance of our DED model.

Reasoning for Machine Learning and Point-Based Scoring Model

Machine learning algorithms and techniques were used for several purposes. First, tree-based machine learning was applied to categorize continuous variables. Second, Lasso was implemented to select important factors to simplify the model and to reduce overfitting. Third, the models were generated using a training sample and validated with a separate test sample, which enabled estimation of predictive power. This technique is preferred because standard regression modeling and automated variable selection (eg, stepwise selection, pretesting of candidate predictors) can result in overfitting [27,28]. As a result, our model was robust enough to generalize to populations not used during training without overfitting (Figure 2).

Point-based scoring systems are useful for describing the relationship between multiple factors and the risk of the development of a disease [15]. Likewise, using our point-based model, DED can be assessed by summing the points accurately describing an individual with a cutoff of 10 points, indicating high risk for DED. In addition, the node size was determined by its point, and interrelatedness of DED risk factors was interrogated. Because DED was predicted by the sum of points, larger nodes might be prioritized in evaluating DED.

Interpretation for Indirect Model Factors With Network Analysis

By risk factor and network analyses, significant factors were presumed to be directedly associated with DED, whereas nonsignificant factors might be indirectly associated. Conventionally, nonsignificant factors might have been confounding variables that are related to DED via other significant factors. The network graph showed that nonsignificant factors such as *phosphorus <746 mg/day* (node 7), *current smoker* (node 12), and *cholesterol ≥240 mg/day* (node 11) were connected to significant factors such as *women* (node 1) and *age 54-66y* (node 8; Figure 3). However, those nonsignificant factors were necessary to maximize the model performance and selected by a machine learning–based Lasso regression. Therefore, they seemed to be included to tune points of other significant factors without a causal effect on DED. For example, *current smoker* (node 12) had a negative effect on node 1 (*women*) because it generally occurred in men rather than women. Smoking has been reported as an inconclusive risk factor for DED, and our study did not suggest smoking as a risk factor [2].

Known Factors in Dry Eye Disease Model

In the network-based analysis in Figure 3, *age 54-66y* (node 8) showed high centrality in the network, which means that it has more connections (strength), it is closer to other nodes (closeness), or makes connections between other nodes (betweenness). This high-centrality node exists at the center of the network and acts as hubs that connect disparate nodes [18]. In contrast, *omega-3 intake* (node 13) and *other ocular surgeries* (node 6) were independent nodes with low centrality.

In the previous study with KNHANES 2010-2011 by Ahn et al [3], 50- to 59-year-old and 60- to 69-year-old groups are presented as risk factors, which are in agreement with our age factor of 54 to 66 years. Other risk factors suggested (women, extreme stress, cataract surgery, refractive surgery, other ocular surgery) were also picked up in our model except for thyroid disease and educational level [3]. Thyroid disease is a possible risk factor, and the previous study argues an ambiguous link between thyroid disease and DED [2,3]. The difference between our work and the previous study can come from different definitions of DED because we used both the diagnosis and symptoms to classify an individual as having DED, whereas the previous study used the criteria of having either the diagnosis or symptoms [3].

Female sex is consistently associated with DED throughout the studies, but the prevalence of DED is considerably variable in these studies with respect to sex and age [2]. Stress has been associated with DED as a trigger or an immune response modulator [2,3]. Ocular surgery can cause DED in various ways, for example, the exposure to strong light of the microscope during the surgery, use of anesthetic or postoperative eyedrops, and the corneal nerve damage [3]. Specifically, refractive surgery leads to neuropathic dry eye by sensory nerve damage, decreased tear secretion, and induced neurogenic inflammation [2].

New Factors in Dry Eye Disease Model

Depression, rhinitis, lipid-lowering medication, and omega-3 intake were new DED-associated factors in our model that were added to previously reported factors of KNHASES [3]. Those factors have not been evaluated in the previous KNHASES study on DED [3]. Depression (node 3) was positively connected to node 1 (women), and a close association between depression and DED in women has been reported [7]. Depression is more prevalent in patients with DED partly because of somatization and perceptional changes in ocular discomfort [29]. In addition, depression was serially connected to other risk factors (Figure 3), such as female sex, stress, and rhinitis, which may be utilized for DED risk evaluation and control because positively connected serial factors can occur together with possible causalities. For rhinitis, allergic rhinitis was reported to be significantly associated with DED, and inflammation is related to both [30,31]. Notably, rhinitis was a clinically reliable factor because it was diagnosed by physician's examination.

Other serial risk factors were *age 54-66y* (node 8), *lipid-lowering medication* (node 10), and cataract surgery (node 4). Dyslipidemia and its treatment might be an issue for the 54- to 66-year-old group, which could explain the negative connection between node 8 and node 11 (*cholesterol >240 mg/day*). Dyslipidemia has been suggested to induce MGD, a major cause of DED [5,32]. However, oral statin therapy, not hypercholesterolemia, were recently reported to be associated

with the symptoms of DED [33]. Interestingly, sterols have been reported to reduce cataract severity [34,35], and cholesterol metabolism might be linked to cataract formation [36].

The results of randomized controlled trials for DED treatment effect of omega-3 have been inconsistent, and larger studies suggest no statistically significant improvement compared with placebo [37,38]. Nonetheless, omega-3 has been commonly used to treat DED in the clinic because essential fatty acids, including omega-3, display anti-inflammatory properties [39], enhance the lipid layer of the tear film, and improve tear secretion while lacking association with substantial side effects [2]. However, it remains a problem that there is no consensus on the dose of supplementation, and our study suggested that 1000 to 1500 mg daily intake of omega-3 (for 2100 kcal average calorie intake) helped to prevent DED. It was noteworthy that *omega-3 intake* might be used to treat DED without possible effects on other factors because it did not have a connection in the network (Figure 3).

Limitations

This study has several limitations. Eye-related factors (blepharitis, lid abnormalities, low blink rate, other ocular surface disease, or conjunctivochalasis) and Sjögren syndrome could not be assessed [40]. In addition, some nutrient factors might have been missed because nutrient intake data were available only for subjects younger than 65 years [9].

Conclusions

In summary, the machine learning–based model to assess the individual risks of DED was successfully created from a large-scale national survey data. With this model, additional DED-associated factors could be suggested, and personalized medical advice was possible using the network graph of the model factors. These approaches allowed integrative understanding of DED and may be applied to other multifactorial diseases.

Acknowledgments

The authors would like to thank Jeremy D Keenan (Department of Ophthalmology, University of California San Francisco, San Francisco, CA) who inspired the authors to incorporate network analysis to show medical relevance. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF-2019R1C1C1007663 and NRF-2017R1E1A1A03070934).

Conflicts of Interest

None declared.

References

1. Craig JP, Nelson JD, Azar DT, Belmonte C, Bron AJ, Chauhan SK, et al. TFOS DEWS II Report Executive Summary. Ocul Surf 2017 Oct;15(4):802-812. [doi: 10.1016/j.jtos.2017.08.003] [Medline: 28797892]
2. Stapleton F, Alves M, Bunya VY, Jalbert I, Lekhanont K, Malet F, et al. TFOS DEWS II Epidemiology Report. Ocul Surf 2017 Jul;15(3):334-365. [doi: 10.1016/j.jtos.2017.05.003] [Medline: 28736337]
3. Ahn JM, Lee SH, Rim THT, Park RJ, Yang HS, Kim TI, Epidemiologic Survey Committee of the Korean Ophthalmological Society. Prevalence of and risk factors associated with dry eye: the Korea National Health and Nutrition Examination Survey 2010-2011. Am J Ophthalmol 2014 Dec;158(6):1205-14.e7. [doi: 10.1016/j.ajo.2014.08.021] [Medline: 25149910]
4. Lee W, Lim SS, Won JU, Roh J, Lee JH, Seok H, et al. The association between sleep duration and dry eye syndrome among Korean adults. Sleep Med 2015 Nov;16(11):1327-1331. [doi: 10.1016/j.sleep.2015.06.021] [Medline: 26498231]
5. Chun YH, Kim HR, Han K, Park YG, Song HJ, Na KS. Total cholesterol and lipoprotein composition are associated with dry eye disease in Korean women. Lipids Health Dis 2013 Jun 5;12:84 [FREE Full text] [doi: 10.1186/1476-511X-12-84] [Medline: 23734839]
6. Chung SH, Myong JP. Are higher blood mercury levels associated with dry eye symptoms in adult Koreans? A population-based cross-sectional study. BMJ Open 2016 Apr 27;6(4):e010985 [FREE Full text] [doi: 10.1136/bmjopen-2015-010985] [Medline: 27121705]
7. Na KS, Han K, Park YG, Na C, Joo C. Depression, stress, quality of life, and dry eye disease in Korean women: a population-based study. Cornea 2015 Jul;34(7):733-738. [doi: 10.1097/ICO.0000000000000464] [Medline: 26002151]
8. Consejo A, Melcer T, Rozema JJ. Introduction to Machine Learning for Ophthalmologists. Semin Ophthalmol 2018 Nov 30:1-23. [doi: 10.1080/08820538.2018.1551496] [Medline: 30500302]
9. Korea Centers for Disease Control and Prevention. Korea National Health & Nutrition Examination Survey URL: https://knhanes.cdc.go.kr/knhanes/eng/index.do [accessed 2019-12-31]
10. Kweon S, Kim Y, Jang MJ, Kim Y, Kim K, Choi S, et al. Data resource profile: the Korea National Health and Nutrition Examination Survey (KNHANES). Int J Epidemiol 2014 Feb;43(1):69-77 [FREE Full text] [doi: 10.1093/ije/dyt228] [Medline: 24585853]
11. Yoon KC, Choi W, Lee HS, Kim SD, Kim SH, Kim CY, et al. An overview of Ophthalmologic survey methodology in the 2008-2015 Korean National Health and Nutrition Examination Surveys. Korean J Ophthalmol 2015 Dec;29(6):359-367 [FREE Full text] [doi: 10.3341/kjo.2015.29.6.359] [Medline: 26635451]

12. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 1972 Jun;18(6):499-502 [FREE Full text] [Medline: 4337382]
13. Padilla O. Merck Manuals. Blood Tests: Normal Values URL: https://www.merckmanuals.com/professional/appendixes/normal-laboratory-values/blood-tests-normal-values#v8508809 [accessed 2019-12-31]
14. Ministry of Health and Welfare, The Korean Nutrition Society. Dietary reference intakes for Koreans 2015. Sejong: Ministry of Health and Welfare; Dec 31, 2015. URL: http://www.mohw.go.kr/react/jb/sjb030301vw.jsp?PAR_MENU_ID=03&MENU_ID=032901&CONT_SEQ=337356
15. Austin PC, Lee DS, D'Agostino RB, Fine JP. Developing points-based risk-scoring systems in the presence of competing risks. Stat Med 2016 Sep 30;35(22):4056-4072 [FREE Full text] [doi: 10.1002/sim.6994] [Medline: 27197622]
16. Epskamp S, Cramer AO, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: network visualizations of relationships in psychometric data. J Stat Soft 2012;48(4):1639-1641. [doi: 10.18637/jss.v048.i04]
17. Pereira-Morales AJ, Adan A, Forero DA. Network analysis of multiple risk factors for mental health in young Colombian adults. J Ment Health 2019 Apr;28(2):153-160. [doi: 10.1080/09638237.2017.1417568] [Medline: 29265896]
18. Robinaugh DJ, Millner AJ, McNally RJ. Identifying highly influential nodes in the complicated grief network. J Abnorm Psychol 2016 Aug;125(6):747-757 [FREE Full text] [doi: 10.1037/abn0000181] [Medline: 27505622]
19. Therneau T, Atkinson B. 2019. rpart: Recursive Partitioning and Regression Trees URL: https://CRAN.R-project.org/package=rpart [accessed 2019-12-31]
20. Kuhn M. 2019. caret: Classification and Regression Training URL: https://CRAN.R-project.org/package=caret [accessed 2019-12-31]
21. Revelle W. 2019. 2019. psych: Procedures for Psychological, Psychometric, and Personality Research URL: https://CRAN.R-project.org/package=psych [accessed 2019-12-31]
22. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):1-22 [FREE Full text] [Medline: 20808728]
23. Lumley T. 2019. survey: Analysis of Complex Survey Samples URL: https://CRAN.R-project.org/package=survey [accessed 2019-12-31]
24. Hocking TD. 2019. WeightedROC: Fast, Weighted ROC Curves URL: https://CRAN.R-project.org/package=WeightedROC [accessed 2019-12-31]
25. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.
26. Pasek J. 2018. weights: Weighting and Weighted Statistics URL: https://CRAN.R-project.org/package=weights [accessed 2019-12-31]
27. McNeish DM. Using Lasso for predictor selection and to assuage overfitting: a method long overlooked in behavioral sciences. Multivariate Behav Res 2015;50(5):471-484. [doi: 10.1080/00273171.2015.1036965] [Medline: 26610247]
28. Babyak MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. Psychosom Med 2004;66(3):411-421. [doi: 10.1097/01.psy.0000127692.23278.a9] [Medline: 15184705]
29. Wan KH, Chen LJ, Young AL. Depression and anxiety in dry eye disease: a systematic review and meta-analysis. Eye (Lond) 2016 Dec;30(12):1558-1567 [FREE Full text] [doi: 10.1038/eye.2016.186] [Medline: 27518547]
30. Yenigun A, Dadaci Z, Sahin GO, Elbay A. Prevalence of allergic rhinitis symptoms and positive skin-prick test results in patients with dry eye. Am J Rhinol Allergy 2016;30(2):e26-e29. [doi: 10.2500/ajra.2016.30.4275] [Medline: 26980382]
31. Yenigun A, Elbay A, Dogan R, Ozturan O, Ozdemir MH. A pilot study investigating the impact of topical nasal steroid spray in allergic rhinitis patients with dry eye. Int Arch Allergy Immunol 2018;176(2):157-162. [doi: 10.1159/000488599] [Medline: 29734186]
32. Rathnakumar K, Ramachandran K, Baba D, Ramesh V, Anebaracy V, Vidhya R, et al. Prevalence of dry eye disease and its association with dyslipidemia. J Basic Clin Physiol Pharmacol 2018 Mar 28;29(2):195-199. [doi: 10.1515/jbcpp-2017-0001] [Medline: 29150990]
33. Ooi KG, Lee MH, Burlutsky G, Gopinath B, Mitchell P, Watson S. Association of dyslipidaemia and oral statin use, and dry eye disease symptoms in the Blue Mountains Eye Study. Clin Exp Ophthalmol 2019 Mar;47(2):187-192. [doi: 10.1111/ceo.13388] [Medline: 30203595]
34. Zhao L, Chen X, Zhu J, Xi Y, Yang X, Hu L, et al. Lanosterol reverses protein aggregation in cataracts. Nature 2015 Jul 30;523(7562):607-611. [doi: 10.1038/nature14650] [Medline: 26200341]
35. Makley LN, McMenimen KA, DeVree BT, Goldman JW, McGlasson BN, Rajagopal P, et al. Pharmacological chaperone for α-crystallin partially restores transparency in cataract models. Science 2015 Nov 6;350(6261):674-677 [FREE Full text] [doi: 10.1126/science.aac9145] [Medline: 26542570]
36. Yamauchi Y, Rogers MA. Sterol Metabolism and Transport in Atherosclerosis and Cancer. Front Endocrinol (Lausanne) 2018;9:509 [FREE Full text] [doi: 10.3389/fendo.2018.00509] [Medline: 30283400]
37. Ton J, Korownyk C. Omega-3 supplements for dry eye. Can Fam Physician 2018 Nov;64(11):826 [FREE Full text] [Medline: 30429179]
38. Dry Eye Assessment and Management Study Research Group, Asbell PA, Maguire MG, Pistilli M, Ying GS, Szczotka-Flynn LB, et al. n-3 Fatty Acid Supplementation for the Treatment of Dry Eye Disease. N Engl J Med 2018 May 3;378(18):1681-1690 [FREE Full text] [doi: 10.1056/NEJMoa1709691] [Medline: 29652551]

39. Serhan CN, Chiang N, van Dyke TE. Resolving inflammation: dual anti-inflammatory and pro-resolution lipid mediators. Nat Rev Immunol 2008 May;8(5):349-361 [FREE Full text] [doi: 10.1038/nri2294] [Medline: 18437155]
40. Clayton JA. Dry Eye. N Engl J Med 2018 Jun 7;378(23):2212-2223. [doi: 10.1056/NEJMra1407936] [Medline: 29874529]

Abbreviations

AUC: area under the curve
DED: dry eye disease
KCDC: Korea Centers for Disease Control and Prevention
KNHANES: Korea National Health and Nutrition Examination Survey
MGD: Meibomian gland dysfunction
NRF: National Research Foundation of Korea
OR: odds ratio
ROC: receiver-operating characteristic

Edited by G Eysenbach; submitted 06.09.19; peer-reviewed by DA Forero, J Bian; comments to author 29.10.19; revised version received 23.11.19; accepted 16.12.19; published 20.02.20.

Please cite as:
Nam SM, Peterson TA, Butte AJ, Seo KY, Han HW
Explanatory Model of Dry Eye Disease Using Health and Nutrition Examinations: Machine Learning and Network-Based Factor Analysis From a National Survey
JMIR Med Inform 2020;8(2):e16153
URL: http://medinform.jmir.org/2020/2/e16153/
doi:10.2196/16153
PMID:32130150

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org