

# A near real-time electronic health record-based COVID-19 surveillance system: An experience from a developing country

Abbas Sheikhtaheri, PhD<sup>1</sup> ,

Seyed Mohammad Tabatabaee Jabali, MD, MPH, PhD<sup>2</sup>,

Ehsan Bitaraf, MD<sup>3</sup> ,

Alireza TehraniYazdi, MD, PhD<sup>3</sup>,

Ali Kabir, MD, MPH, PhD<sup>4</sup>

## Abstract

**Context:** Access to real-time data that provide accurate and timely information about the status and extent of disease spread could assist management of the COVID-19 pandemic and inform decision-making. **Aim:** To demonstrate our experience with regard to implementation of technical and architectural infrastructure for a near real-time electronic health record-based surveillance system for COVID-19 in Iran. **Method:** This COVID-19 surveillance system was developed from hospital information and electronic health record (EHR) systems available in the study hospitals in conjunction with a set of open-source solutions; and designed to integrate data from multiple resources to provide near real-time access to COVID-19 patients' data, as well as a pool of health data for analytical and decision-making purposes. **Outcomes:** Using this surveillance system, we were able to monitor confirmed and suspected cases of COVID-19 in our population and to automatically notify stakeholders. Based on aggregated data collected, this surveillance system was able to facilitate many activities, such as resource allocation for hospitals, including managing bed allocations, providing and distributing equipment and funding, and setting up isolation centres. **Conclusion:** Electronic health record systems and an integrated data analytics infrastructure are effective tools to enable policymakers to make better decisions, and for epidemiologists to conduct improved analyses regarding COVID-19. **Implications:** Improved quality of clinical coding for better case finding, improved quality of health information in data sources, data-sharing agreements, and increased EHR coverage in the population can empower EHR-based COVID-19 surveillance systems.

## Keywords (MeSH)

Public health informatics, public health surveillance, COVID-19, Hospital information systems, electronic health records, clinical coding, health information management,

## Supplementary Keywords

Public health information, National health information network, Dashboard, Disease registry

## Introduction

COVID-19 spread rapidly across many countries of the world, including Iran (Zhu et al., 2020; World Health Organization[WHO], 2020, 2022). According to WHO (2022) data, more than 503 million people worldwide have been infected, with Iran being one of the countries with the highest number of patients and deaths. The incidence of this disease in Iran has continued to increase (with more than 7,200,000 people infected since late February 2020 to date (April 18, 2022) (WHO, 2022).

While all affected countries acted to address this disease, the effectiveness of their actions and decisions (clinical and non-clinical) depended on the accuracy and timeliness of information available (e.g., about the extent of the outbreak;

<sup>1</sup>Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran

<sup>2</sup>Department of Sports Medicine, School of Medicine, Iran University of Medical Sciences, Tehran, Iran

<sup>3</sup>Center for Statistics and Information Technology, Iran University of Medical Sciences, Tehran, Iran

<sup>4</sup>Minimally Invasive Surgery Research Center, Iran University of Medical Sciences, Tehran, Iran

Accepted for publication May 15, 2022

## Corresponding author:

Ali Kabir, Minimally Invasive Surgery Research Center, Iran University of Medical Sciences, Maziar Mansouri St., Niayesh St., Sattarkhan Ave, Tehran, Iran, 1449614535.

Email: [kabir.a@iums.ac.ir](mailto:kabir.a@iums.ac.ir)

the characteristics of people most at risk). This highlights the importance of not only accurate health information for the management of the pandemic but, as [Sitting and Singh \(2020\)](#) pointed out, for effective health information management infrastructure to be set up or strengthened to assist timely and effective decision-making, to facilitate the gathering together of appropriate information about the disease outbreak for identification of hotspots, allocation of resources, and planning interventions; and for the evaluation of the effectiveness of these interventions ([Ulahannan et al., 2020](#); [Wissel et al., 2020](#)). With this in mind, many researchers around the world have developed information infrastructure to gather, analyse, and distribute information about COVID-19 in disease registries or data-sharing networks ([Brat et al., 2020](#); [Dagliati et al., 2021](#); [Dong et al., 2020](#); [Madhavan et al., 2021](#); [Melissa et al., 2020](#); [Wissel et al., 2020](#); [Zarei et al., 2021](#)). These data-sharing initiatives depended on different data sources with different data formats and low rates of data standardisation ([Dagliati et al., 2021](#); [Madhavan et al., 2021](#)), and developing COVID-19 registries or data-sharing networks is also costly. On the other hand, electronic health record (EHR) systems, “an electronic record of health-related information on an individual that conforms to nationally recognised interoperability standards and that can be created, managed and consulted by authorised clinicians and staff across more than one healthcare organisation” (National Alliance for Health Information Technology, cited by [Wager et al., 2013](#): 136), can offer real potential to track COVID-19 pandemic status.

The purpose of this article is to describe the technical and architectural infrastructure of the completely free and open-source EHR-based COVID-19 surveillance system that we have developed in Iran, and the lessons we have learned from this experience.

### Study context

In Iran, a number of software vendors have implemented a wide variety of health information systems, including, for example, hospital information systems, primary care information systems, and laboratory information systems. The Iranian Ministry of Health and Medical Education (MOHME) has also established and implemented an integrated infrastructure for health information exchange ([Bitaraf et al., 2021b](#); [2021c](#)). To electronically collect and exchange health data, any point of care unit (POCU) is now equipped with an information system (e.g., medical laboratories use laboratory information systems (LIS); hospitals use hospital information systems (HIS)). However, these systems have been developed within the private sector, and their diversity has created challenges for data integration and interoperability.

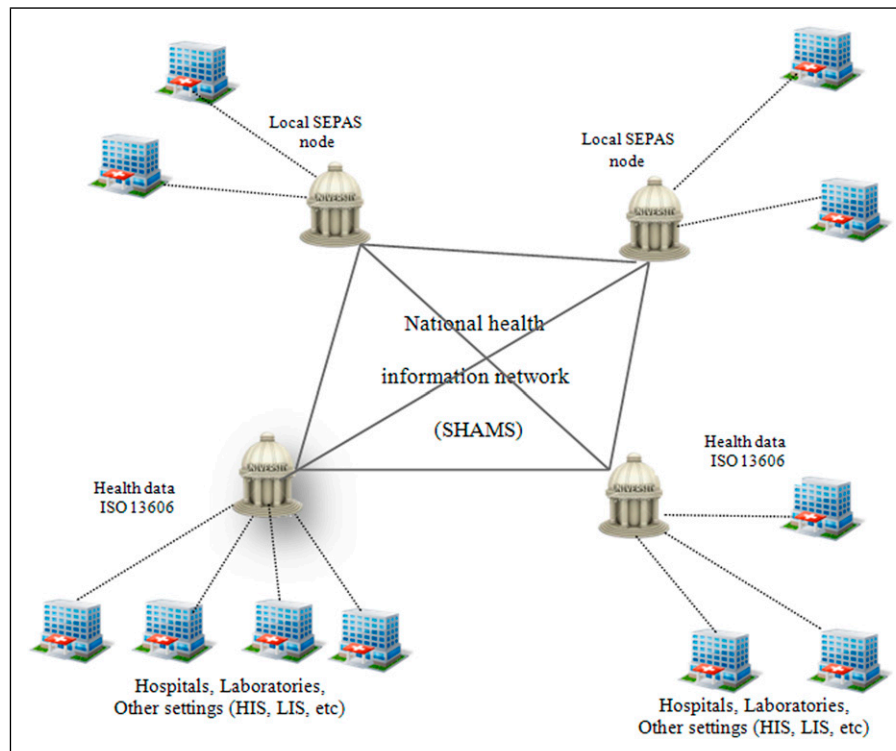
The MOHME developed a set of standards and services through which POCUs can exchange health information, and has also implemented a National Integrated Electronic Health Record System, known locally as SEPAS ([Bitaraf et al., 2021b](#); [Riazi et al., 2010, 2014](#)). In Iran, EHR is defined as the electronic collection of information related to the health of Iranian citizens, from birth to death, which is longitudinally maintained. Based on this definition, hospital

and non-hospital data should be covered by EHRs ([Riazi et al., 2010](#)).

Hospital information systems can also exchange electronic health data, such as billing summaries, laboratory test results, prescriptions, diagnoses, and procedures through the SEPAS, which has a distributed architecture based on ISO 13606 and inspired by OpenEHR architecture ([Bitaraf et al., 2021b](#); [Riazi et al., 2010, 2014](#)). The software infrastructure instances are installed in different locations called SEPAS nodes, which are hosted by each medical university in the country. Hospitals operate under the supervision of medical universities, and they exchange data through their corresponding SEPAS nodes ([Riazi et al., 2014](#)). This infrastructure is equipped with a record locator function. In Iran, the unique patient identifier is the national identification number and all POCUs in the country are required to use this ID, which means the infrastructure can identify patients and integrate their records. To guarantee high bandwidth and secure communication channels, the health information systems and SEPAS interoperate in a national private health information network, locally called the SHAMS network. [Figure 1](#) illustrates the conceptual schema. In addition, outpatient and primary POCUs use electronic primary care health records nationally. The most important electronic primary EHRs (developed by the private vendors) are referred to locally as SIB and SINA. They cover most of the primary POCUs in the country.

According to MOHME guidelines, COVID-19 patients are classified into two categories: confirmed cases (based on laboratory results); and suspected cases (based on signs and symptoms before the laboratory result is confirmed) ([Iranian Ministry of Health and Medical Education, 2020](#)). In Iran, hospitals and their laboratories provide healthcare services to these patients, including PCR tests, and the relevant information is stored in HISs. In addition, some laboratories collect PCR tests. These POCUs store the patient and laboratory test-related data in LISs. Moreover, each primary care POCU performs tests and outpatient care, and stores the related data in primary care information systems (SINA or SIB). SEPAS acts as the health-related data integrator. In other words, this infrastructure exchanges and stores the health-related data for each individual in an integrated manner. As illustrated in [Figure 1](#), POCUs transfer data electronically through their corresponding SEPAS nodes. It is noteworthy that these data are as original as the data generated and used at POCUs, which means the systems can communicate and transfer necessary data without any additional data entry and workforce input ([Bitaraf et al., 2021b](#)). Despite these available health information infrastructures, the systems have not been widely used for public health information management and disease surveillance. For instance, during the COVID-19 pandemic, a new reporting system was launched by the MOHME, referred to locally as MCMC, to report COVID-19 cases by healthcare providers, which faces problems such as data quality, underreporting, high cost, and manpower issues ([Zarei et al., 2021](#)).

As one of the first initiatives, the Iran University of Medical Sciences (IUMS), in collaboration with the MOHME, used this information infrastructure for near real-time surveillance of the COVID-19 outbreak in the



**Figure 1.** Iranian health information network.

population covered by this university. IUMS is one of three medical science universities located in Tehran, which has a population of more than 13 million people. Through its affiliated hospitals and primary care settings, IUMS covers about 5.4 million people (being the most populated area in the country). There are 18 (11 teaching and 7 non-teaching) hospitals and 499 primary care centres operating under the supervision of IUMS.

We established the IUMS COVID-19 Incident Command Center (IUMS-CICC) in the university at the beginning of the outbreak. The IUMS-CICC developed a multidisciplinary team of clinical, epidemiology, health information technology, and some experts in health information management to develop the COVID-19 information infrastructure for the IUMS-CICC. The IUMS hosts an instance of a SEPAS node, which receives large numbers of hospital transactions (over 300,000 transactions a day, approximately). This node integrates all health-related data for each individual (citizen) from several POCUs (hospitals, laboratories, primary healthcare centres) and maintains the history of patient encounters longitudinally. Furthermore, the primary care centres covered by our system used a primary care EHR system (SINA) to record healthcare services, and our COVID-19 surveillance and data analysis architecture implemented was based on the IUMS SEPAS node and SINA.

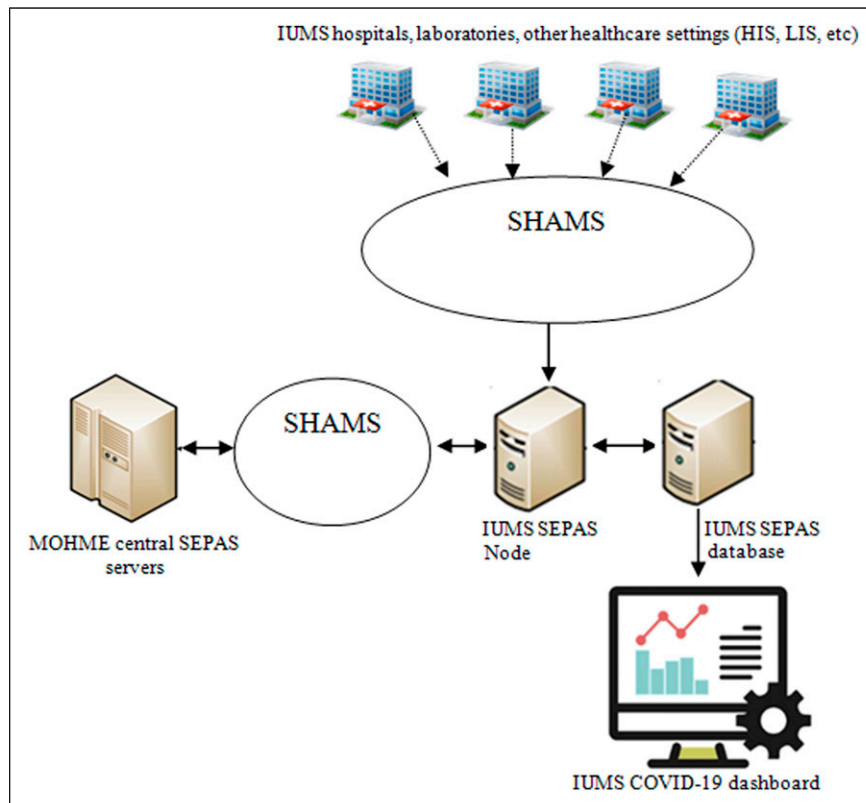
Our healthcare facilities exchanged health data through the SEPAS and SHAMS infrastructure. Patients' data were sent to the IUMS SEPAS node according to ISO 13606 standard and data exchange guidelines compiled by MOHME. To develop the COVID-19 surveillance system, it was necessary to extract patients' data from this infrastructure. Following the WHO (2020) announcement of emergency U07.1 and U07.2 ICD-10 codes for confirmed (laboratory-confirmed) and suspected (clinically diagnosed)

COVID-19 patients, the MOHME developed a coding and documentation guideline for hospitals to record these data in HISs and any other health information systems. Therefore, our case-finding algorithms have been mainly based on these diagnosis codes. Figure 2 illustrates the technical infrastructure of the system from the software point of view.

### Review of literature on similar cases

Although there are many barriers and challenges to using available EHR data for disease surveillance (Aliabadi et al., 2020; Madhavan et al., 2021), EHR systems can provide timely, standardised, and low-cost data for health surveillance systems and provide public health analytics without gathering duplicate data (Kukafka et al., 2007; Institute of Medicine, 2011); and the use of EHR for public health surveillance has been studied by researchers from all over the world (Buck et al., 2012; Burkom et al., 2021; Izadi et al., 2013; Joshi et al., 2017; Lazarus et al., 2009; Newton-Dame et al., 2016; Perlman et al., 2017; Shaban-Nejad et al., 2017; Vogel et al., 2014; Willis et al., 2019).

Researchers at Johns Hopkins University (Dong et al., 2020) developed an online interactive dashboard to visualise the location and number of confirmed and recovered patients and deaths for affected countries, in real-time. The information was based on official aggregated data reported from various countries, not basically EHRs. In another study in the United States (US) (Wissel et al., 2020), researchers compared three COVID-19 dashboards, including The New York Times (NYT) COVID-19 data, Johns Hopkins University COVID-19 data, and the COVID Tracking Project data, and developed a system for visualising COVID-19 data in the US. This system, COVID-19 Watcher, checks the information of NYT and COVID Tracking Project every hour,



**Figure 2.** The infrastructure of the COVID-19 surveillance system in Iran University of Medical Sciences.

downloads new data, and after quality control, visualises it in a dashboard. In a third study (Ulahannan et al., 2020), the authors reported a dashboard to visualise real-time COVID-19 information for the public. Authors used data extracted from daily bulletins released by official authorities and various news outlets. Developers also collected data from public volunteers through social network channels. Following data cleaning and verification, they developed a live structured dataset for real-time data analysis and visualisation.

The University of California, San Diego Health (Wissel et al., 2020), implemented various technologies to deal with COVID-19, including EHR-based reporting and analytics tools. This system generates a variety of reports that are automatically available to decision-makers via mobile devices. These reports are used to make decisions about expanding test capacity, isolating patients, and monitoring compliance with screening procedures. Brakefield et al. (2020, 2021) also reported Urban Public Health Observatory system in which different data sources including EHRs were used for COVID-19 data analytics and surveillance. In addition, many other researchers have suggested EHR data-sharing for collaborative research or public health (Brat et al., 2020; Dagliati et al., 2021; Madhavan et al., 2021; Melissa et al., 2020). Therefore, available COVID-19 patients' data usually stored in HISs, EHRs, and LISs can be valuable sources for COVID-19 surveillance.

## Case study

### Data sources

Several data sources have contributed to our surveillance and exchange data on this platform (mainly through SEPAS),

including hospital information systems from 31 hospitals and their laboratory information systems. These systems differ from those developed by private vendors. Furthermore, 499 primary care centres have also contributed to this surveillance through the primary healthcare information system (SINA) used in IUMS.

### Surveillance system architecture

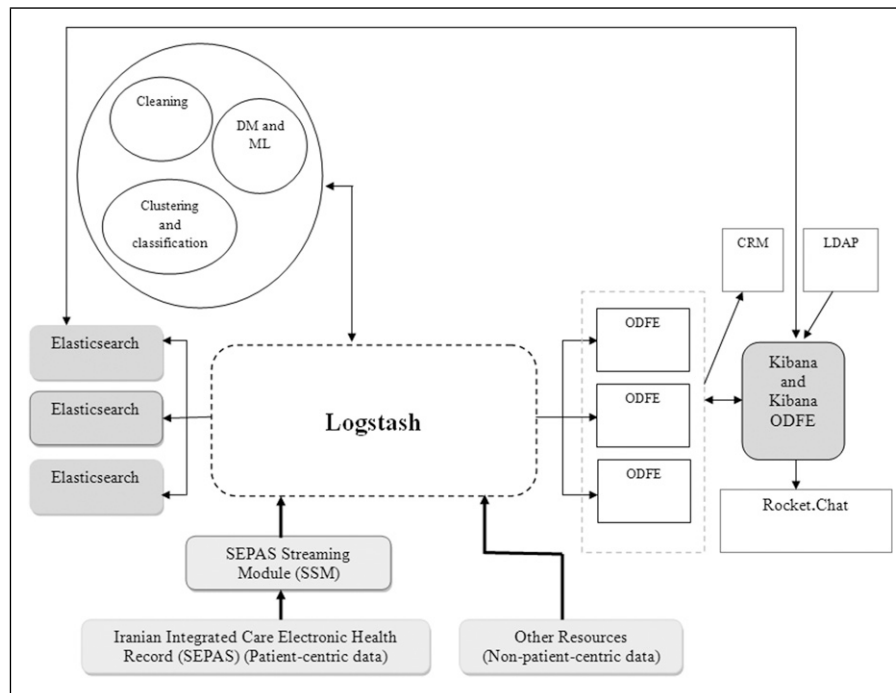
To review the current infrastructure, we conducted a rapid situational analysis that included MOHME standards and recommendations from the literature regarding EHR-based surveillance systems. After several group discussions, we finally decided to use open-source solutions to develop our COVID-19 data pipeline to identify cases, extract data, and distribute relevant information. Figure 3 shows the conceptual architecture of the COVID-19 surveillance system. This technical architecture is composed of several components, as described below.

### The SEPAS Streaming Module

To use the EHR data and extract data from the SEPAS node, there are two general approaches:

- **Offline Approach:** in this approach, the data extraction process starts after the data are stored in the SEPAS node. This process is secure using a hardware token for authentication, authorisation, and a digital signature of the requests and responses. In this approach, the authentication/authorisation process must take place each time an electronic record or patient encounter is





**Figure 3.** The technical architecture of the COVID-19 surveillance system in Iran University of Medical Sciences.

accessed. This is a useful and secure approach but needs plenty of time to extract data, and is also dependent on an expert user to start the data extraction process.

- **Online Approach:** in this approach, the data are processed and anonymised, and used for further applications based on predefined rules. Asymmetric key pairs systematically perform the authentication, authorisation, and digital signature of requests and responses, thus guaranteeing the security of the process. In addition, data are processed at the same time as they are stored in the SEPAS node, resolving the problem of dependency on an expert to initiate the data extraction process. The online approach occurs via the SEPAS Streaming Module (SSM) developed by the MOHME to facilitate utilisation of data through the SEPAS data stream, by which proper data marts or data warehouses can receive and store the health data.

### Elasticsearch and Logstash

The SSM is responsible only for the online and automated EHR data extraction process. This means that it is not responsible for data storage in secondary data warehouses or even processing the data for analytical purposes. To store the data into a data warehouse, there is a need to design and implement proper data warehouse infrastructure as well as proper data processing entities and tools. In this regard, we used Elasticsearch, which is a free open-source distributed document-oriented search and analytics engine for all types of structured and unstructured queries (Barr, 2019; Brasetvik, 2020). We developed our COVID-19 data marts based on Elasticsearch. In our infrastructure, the SSM receives SEPAS EHR extracts, prepares, and passes these data to Elasticsearch and Logstash.

Logstash is a free, open-source data collection engine with a scalable data pipeline that provides features such as powerful

search and Kibana dashboard for data visualisation. Logstash can also provide facilities for generating geo-coordinates and longitude and latitude as well as data anonymisation. Additionally, it is a pluggable pipeline and we can add many HTTP plugins (Logstash, 2020). In our infrastructure, we have converted the text-based patients' addresses to longitude and latitude to be visualised on the map. We also added machine learning and data mining services (R and Python) to this infrastructure.

### Open Distro for Elasticsearch

Elasticsearch has some limitations, including no free features for authentication or authorisation (Brasetvik, 2020), or other functionalities (Barr, 2019). For example, some security services, online data analytics, machine learning services, email alerts and notifications via SMS and social networks, and JDBC connections to SQL databases are not free. To overcome these limitations, we used Open Distro for Elasticsearch (ODFE). ODFE is a completely free and open-source version for Elasticsearch and Kibana supported by Amazon Web Services (AWS). It provides advanced security features (such as Lightweight Directory Access Protocol (LDAP) for role-based access control and user authentication), audit logging, event monitoring and alerting (to monitor data and generate alerts and automatically notify stakeholders based on predefined conditions), performance analysis, and free SQL query support and JDBC connection to SQL databases (Barr, 2019). As shown in Figure 3, in our COVID-19 information infrastructure, we send data to ODFE and visualise these data using Kibana.

### Customer Relationship Management and Rocket.Chat

We used ODFE to connect other systems to this infrastructure. For example, we have currently connected our

customer relationship management (CRM) system to this infrastructure to communicate with patients and follow them up. Furthermore, we developed a dedicated internal COVID-19 social network based on the free open-source Rocket.Chat social media platform (RocketChat, 2020) to automatically send information and alerts to stakeholders and policymakers at the university level.

### System features and functionalities

According to the Ministry of Health guideline (Iranian Ministry of Health and Medical Education, 2020) for COVID-19 patients, all patients with clinical diagnoses (based on the related signs and symptoms) are considered as suspected cases. The POCUs register the reason for the encounter of these patients as suspected COVID-19 (U07.2 ICD-10 code) at the time of admission. The relevant health information systems in these POCUs automatically transfer these data through SEPAS nodes. After successful storage of the patient encounter in SEPAS nodes, it sends an instance of these records to SSM. SSM authenticates the sender and receiver, anonymises the patient record, and sends the record to Logstash. Thus, the data extraction process is performed instantly after patient admission and the policymakers have real-time online access to information on the status of COVID-19 outbreak in the covered population. When the patients are registered as confirmed cases (U07.1 ICD-10 code) in data sources (following laboratory test confirmation) and the information system registers the confirmed diagnosis, our system detects them as confirmed cases. This system provides near real-time access to several useful types of information for the policymakers including the number of confirmed cases, number of suspected cases, distribution of admitted patients by hospitals/centres, number of discharges, number of recovered patients and deaths, the time trend of patients' admission and discharge, patients' costs, geo-map visualisations of patients based on place of residence, and the number of laboratory tests (positive or negatives) (Supplementary Figures S1–S5)

### Discussion

Figure S1, online supplement shows a sample screenshot of the number of suspected, confirmed, and recovered patients as well as the number of patients referred to other centres and discharged against medical advice. The bar chart also shows the number of cases in each hospital. Users can filter the time and change this visualisation based on different timeframes. Figure S2, online supplement shows the time trends of patients' admission per hospital during January 2022. Users can filter the hospitals and change the time. Figures S3 and S4, online supplement show the heat map and point map of the distribution of hospitalised patients. Figure S5, online supplement also shows our dashboard for patients identified from primary care centres. Other features include a dedicated COVID-19 social network for university officials, policymakers, and for the university health system to receive alerts and notifications automatically through this social network via any smartphone. This infrastructure also enables us to link these data to patients' cell phone-based location data to monitor transportation and travels of patients in the city or in

the country to monitor the probable spread of the disease. For this feature, our system sends patients' cell phone numbers to the Iranian Telecommunications Company and receives transportation data using a hash algorithm. The visualisation of output on the map is anonymous. These maps provide the movement and travel patterns to enable policymakers to develop and decide on possible travel restrictions.

The CRM enables our staff to call patients and to follow up to gather additional data, such as patients' reported outcomes. In addition, the machine learning platform enables researchers to analyse the data. For example, many university researchers are working on such data to predict future trends of the disease or to develop models to predict length of stay, ICU admissions, and deaths, or to investigate treatments (Yahyavi et al., 2021; Bitaraf et al., 2021a).

### Current status of the system

As of February 7, 2022, there were 202,528 records related to suspected cases (patients), and 72,872 electronic records for confirmed cases. Total mortality among confirmed hospitalised patients was 12.37% ( $n = 9019$ ) (see Box 1). There were also 89,222 positive cases based on laboratory tests reported by primary care centres.

*What can be learnt from this case?* We developed a near real-time surveillance system based on existing HIS and EHR infrastructure to monitor the COVID-19 pandemic in our covered population, and a dedicated social network to automatically send notifications and alerts to stakeholders. Our surveillance system enabled us to monitor the number of patients in the covered population and to examine the hospital and primary care centre activities, including the status of the general and ICU beds. Based on the information obtained from this system, we have made several decisions to equip hospitals and allocate funds to strengthen healthcare services and set up isolation centres. Other healthcare organisations and universities around the world have also launched similar online monitoring programmes; however, many of their COVID-19 dashboards use official statistics related to confirmed or suspected cases to visualise these data and do not reuse patients' EHR data for COVID-19 surveillance. Using this surveillance system, IUMS and the regional COVID-19 command centre made different decisions regarding the disease. Despite these benefits, the implementation of this EHR-based surveillance faces some challenges.

Different methods, definitions, and standards used for data collection in hospitals (Atreja et al., 2008; Dagiati et al., 2021; Krusina et al., 2020; Tomines et al., 2013) create some problems for developing case-finding algorithms to identify and extract relevant data. The unknown nature of this disease and frequent changes to definitions (Madhavan et al., 2021) at the beginning of this pandemic challenged us to identify cases. In the early stages of launching our system, we identified patients with some symptoms and some ICD-10 codes related to respiratory diseases and pneumonia that resulted in problems in confirming the cases. However, after the WHO announcement of U07 emergency codes and the MOHME COVID-19 documentation and coding guidelines, we currently use these codes for case-finding.

**Box 1.** Distribution of COVID-19 patients in hospitals connected to IUMS COVID-19 surveillance system

Hospital code	Suspected patients <sup>a</sup>	Confirmed patients	ICU patients	Death
H1	34,839	16,946	3716	1677
H2	34,450	4915	843	865
H3	28,350	8133	1603	1378
H4	23,418	7699	0	1731
H5	18,814	1034	643	128
H6	13,137	8990	804	730
H7	9383	3603	1173	737
H8	8072	1	0	0
H9	7220	2100	576	518
H10	5418	274	51	7
H11	5140	2756	738	230
H12	4242	691	9	47
H13	3638	2017	259	33
H14	3051	44	0	1
H15	1564	2036	0	3
H16	1058	156	5	43
H17	561	1	0	0
H18	173	1299	402	146
H19	0	3544	0	363
H20	0	2982	1224	226
H21	0	1545	155	56
H22	0	1533	49	19
H23	0	165	90	29
H24	0	145	0	0
H25	0	142	97	29
H26	0	49	16	12
H27	0	38	0	5
H28	0	13	10	6
H29	0	9	2	0
H30	0	8	0	0
H31	0	4	0	0
Total	202,528	72,872	12,465	9019

<sup>a</sup>All admitted cases with an admission diagnosis of COVID-19.

The quality of EHR data (Kruse et al., 2018; Newton-Dame et al., 2016) is also important. Many EHR-based surveillance systems mainly rely on ICD codes for case finding, and studies indicate that coding errors and timeliness (Farzandipour and Sheikhtaheri, 2009; Handley and Emsley, 2020) may adversely affect the estimation for the incidence of disease, and timeliness of surveillance. In our system, although we can detect suspected hospitalised patients at the time of admission and patients with positive test results at primary care centres immediately after their test results are available, delays in detecting confirmed hospitalised patients may still occur because the coding process is completed after discharge.

There have been few studies that have investigated coding errors in COVID-19 cases. For example, one study found a high sensitivity (94.9%) and positive predictive value (81.2% [95% CI, 80.1%–82.2%]) of the COVID-19-specific code, U07.1 (Cocoros et al., 2021). In our study, we assumed we could detect most of the COVID-19 patients. At the beginning of launching the system, before introducing COVID-19 ICD-10 codes, we regularly double-checked the information obtained from this system with hospital health information managers to ensure data accuracy.

We estimate that our system can detect COVID-19 in most patients. As at February 7, 2022, we had 72,872 and 89,222 confirmed cases from hospitals and primary care centres. Based on our covered population and the officially announced number of cases (World Health Organization, 2020), we estimate that the number of cases in our system was fewer than the actual number of cases in this population. However, this was due to our system coverage, not the detection algorithms. One of the challenges of this method of disease surveillance frequently cited in the literature is the non-complete coverage of the covered population in EHRs (Klompas et al., 2017; McVeigh et al., 2013; Namulanda et al., 2018). All hospitals and primary healthcare centres affiliated with IUMS currently exchange data with our infrastructure. In our geographical area, there are other private hospitals and hospitals affiliated with military or social security organisations. During the last 2 years, many non-affiliated hospitals have also exchanged data with our system; these centres cover 100% of our covered population; however, there are still some hospitals that are not connected to our system. Some patients receive healthcare in these hospitals and reduce our surveillance coverage. Although these patients may be covered by traditional reporting system, increasing the coverage of our

system requires partnership, cooperation, and data-sharing agreements (Krusina et al., 2020; Melissa et al., 2020) and we are currently taking steps to connect other hospitals to our infrastructure.

In EHRs, social and risk factors are often poorly recorded and, consequently, EHR-based surveillance systems are limited regarding these types of data (Elliott et al., 2012; Willis et al., 2019). Our system is similarly limited. In addition, although we have access to signs, symptoms, medications, and healthcare costs of COVID-19 patients, there are few clinical details available in the hospital information systems, and subsequently in their EHRs. Therefore, our system mainly provides the data needed for executive decision-making. Adding clinical documentation templates to health information systems and developing clinical documentation guidelines (Reeves et al., 2020) would strengthen our clinical data analytics platform.

## Conclusion

We have described our public health information infrastructure based on the available infrastructure in Iran for real-time COVID-19 surveillance. Currently, this infrastructure is available at the Iran University of Medical Sciences and covers the corresponding population for this university. This system provides access to the aggregated and individual administrative and clinical data for policymakers and researchers, and enables policymakers to monitor COVID-19 confirmed and suspected cases in our population. Additionally, researchers and epidemiologists can use this infrastructure to reuse EHR-based COVID-19 data to conduct clinical and epidemiological studies. We are currently working with the MOHME to receive ethical and regulatory approval and to develop a data-sharing agreement to expand this surveillance system to the rest of the country. In addition, our platform may be useful for other diseases and future pandemics.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.



## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## Ethical approval

This work received ethical approval from the Research Ethics Committee at the Iran University of Medical Sciences (IR.IUMS.REC.1399.884).

## ORCID iDs

Abbas Sheikhtaheri  <https://orcid.org/0000-0002-6879-5415>  
Ehsan Bitaraf  <https://orcid.org/0000-0002-6588-7349>

## Supplemental Material

Supplemental material for this article is available online.

## References

- Aliabadi A, Sheikhtaheri A and Ansari H (2020) EHR-based Disease surveillance systems: a systematic literature review on challenges and solutions. *Journal of American Medical Informatics Association* 27(12): 1977–1986.
- Atreja A, Gordon SM, Pollock DA, et al. (2008) Opportunities and challenges in utilizing electronic health records for infection surveillance, prevention, and control. *American Journal of Infection Control* 36(3): S37–S46.
- Barr J (2019) *Open Distro for Elasticsearch*. Available at: <https://aws.amazon.com/blogs/aws/new-open-distro-for-elasticsearch/> (accessed 06 May 2020).
- Bitaraf E, Ahmadi S A Y, Gandomi-Mohammadabadi A, et al. (2021a) Effects of immune system-related medications on COVID-19 outcome in a cohort of Iranian patients: preliminary report of a data mining study. *Journal of Immunology Research* 9934134: 1–21. DOI: [10.1155/2021/9934134](https://doi.org/10.1155/2021/9934134)
- Bitaraf E, Jafarpour M, Jami V, et al. (2021b) The Iranian integrated care electronic health record. *Studies in Health Technology and Informatics* 281: 654–658.
- Bitaraf E, Sarani Rad F, Jafarpour M, et al. (2021c) DITAS: an integrated gateway to e-health communications. *Studies in Health Technology and Informatics* 281: 452–456.
- Brakefield WS, Ammar N, Olusanya O, et al. (2020) Implementing an urban public health observatory for (near) real-time surveillance for the COVID-19 pandemic. *Studies in Health Technology and Informatics* 275: 22–26.
- Brakefield W S, Ammar N, Olusanya O A, et al. (2021) An Urban population health observatory system to support covid-19 pandemic preparedness, response, and management: design and development study. *JMIR Public Health and Surveillance* 7(6): e28269.
- Brasetvik A (2020) *Elasticsearch as a NoSQL Database*. Available at: <https://www.elastic.co/blog/found-elasticsearch-as-nosql> (accessed 05 May 2020).
- Brat GA, Weber GM, Gehlenborg N, et al. (2020) International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. *npj Digital Medicine* 3: 109. DOI: [10.1038/s41746-020-00308-0](https://doi.org/10.1038/s41746-020-00308-0)
- Buck MD, Anane S, Taverna J, et al. (2012) The Hub population health system: distributed ad hoc queries and alerts. *Journal of American Medical Informatics Association* 19(e1): 46–50.
- Burkom H, Loschen W, Wojcik R, et al. (2021) Electronic surveillance system for the early notification of community-based epidemics (ESSENCE): overview, components, and public health applications. *JMIR Public Health and Surveillance* 7(6): e26303.
- Cocoros N M, Fuller C C, Adimadhyam S, et al. (2021) A COVID-19-ready public health surveillance system: the food and drug administration's sentinel system. *Pharmacoepidemiology and Drug Safety* 30: 827–837.
- Dagliati A, Malovini A, Tibollo V, et al. (2021) Health informatics and EHR to support clinical research in the COVID-19 pandemic: an overview. *Briefings in Bioinformatics* 22(2): 812–822.
- Dong E, Du H and Gardner L (2020) An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* 20(5): 533–534.
- Elliott AF, Davidson A, Lum F, et al. (2012) Use of electronic health records and administrative data for public health



- surveillance of eye health and vision-related conditions in the United States. *American Journal of Ophthalmology* 154(6 Suppl): S63–S70.
- Farzandipour M and Sheikhtaheri A (2009) Evaluation of factors influencing accuracy of principal procedure coding based on ICD-9-CM: An Iranian study. *Perspectives in Health Information Management* 6(5): 5, Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2682663> (accessed 06 July 2020).
- Handley JD and Emsley HC (2020) Validation of ICD-10 codes shows intracranial venous thrombosis incidence to be higher than previously reported. *Health Information Management Journal* 49(1): 58–61.
- Institute of Medicine (2011) Surveillance—data sources and indicators. *A Nationwide Framework for Surveillance of Cardiovascular and Chronic Lung Diseases*. Washington, DC: The National Academies Press, p. 12. DOI: [10.17226/13145](https://doi.org/10.17226/13145).
- Iranian Ministry of Health and Medical Education (2020) *Guideline for Diagnosis and Treatment of COVID-19 for Outpatient and Inpatient Services*. Tehran: Ministry of Health and Medical Education.
- Izadi MT, Shaban-Nejad A, Okhmatovskaia A, et al. (2013) Population health record: an informatics infrastructure for management, integration, and analysis of large scale population health data. In: *Proceedings Expanding the Boundaries of Health Informatics Using Artificial Intelligence: Papers from the AAAI 2013 Workshop*. Washington, USA, July 2013.
- Joshi A, Thorpe L and Waldron L (2017) *Population Health Informatics*. Sudbury: Jones & Bartlett Learning.
- Klompas M, Cocoros NM, Menchaca JT, et al. (2017) State and local chronic disease surveillance using electronic health record systems. *American Journal of Public Health* 107(9): 1406–1412.
- Kruse CS, Stein A, Thomas H, et al. (2018) The use of electronic health records to support population health: a systematic review of the literature. *Journal of Medical Systems* 42(11): 214.
- Krusina A, Chen O, Otero Varela L, et al. (2020) Developing a data integrated COVID-19 tracking system for decision-making and public use. *International Journal of Population Data Science* 5(4): 1389. DOI: [10.23889/ijpds.v5i1.1389](https://doi.org/10.23889/ijpds.v5i1.1389)
- Kukafka R, Ancker JS, Chan C, et al. (2007) Redesigning electronic health record systems to support public health. *Journal of Biomedical Informatics* 40(4): 398–409.
- Lazarus R, Klompas M, Campion FX, et al. (2009) Electronic support for public health: validated case finding and reporting for notifiable diseases using electronic medical data. *Journal of American Medical Informatics Association* 16(1): 18–24.
- Logstash (2020). Introduction to Logstash. Available at: <https://www.elastic.co/guide/en/logstash/current/introduction.html> (accessed 05 May 2020).
- Madhavan S, Bastarache L, Brown JS, et al. (2021) Use of electronic health records to support a public health response to the COVID-19 pandemic in the United States: a perspective from 15 academic medical centers. *Journal of American Medical Informatics Association* 28(2): 393–401.
- McVeigh K, Newton-Dame R, Perlman S, et al. (2013) *Developing an Electronic Health Record-Based Population Health Surveillance System*. New York: New York City Department of Health and Mental Hygiene.
- Melissa H, Christopher C and Kenneth G (2020) The National COVID Cohort Collaborative (N3C): rationale, design, infrastructure, and deployment. *Journal of American Medical Informatics Association* 28(3): 427–443.
- Namulanda G, Qualters J, Vaidyanathan A, et al. (2018) Electronic health record case studies to advance environmental public health tracking. *Journal of Biomedical Informatics* 79(1): 98–104.
- Newton-Dame R, McVeigh KH, Schreiberstein L, et al. (2016) Design of the New York City MacroScope: innovations in population health surveillance using electronic health records. *EGEMS (Washington, DC)* 4(1): 1265–1265.
- Perlman SE, McVeigh KH, Thorpe LE, et al. (2017) Innovations in population health surveillance: using electronic health records for chronic disease surveillance. *American Journal of Public Health* 107(6): 853–857.
- Reeves JJ, Hollandsworth HM, Torriani FJ, et al. (2020) Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. *Journal of the American Medical Informatics Association* 27(6): 853–859.
- Riazi H, Fathi B and Bitaraf E (2010) *Electronic Health Record System: Concepts, Standards and Development Solutions*. Tehran: Ministry of Health and Medical Education.
- Riazi H, Jafarpour M and Bitaraf E (2014) Towards national ehealth implementation—a comparative study on WHO/ITU national eHealth strategy toolkit in Iran. *Studies in Health Technology and Informatics* 205: 246–250.
- RocketChat (2020). *Rocket.Chat documentation*. Available at: <https://docs.rocket.chat/> (accessed 06 May 2020).
- Shaban-Nejad A, Lavigne M, Okhmatovskaia A, et al. (2017) PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data. *Annals of the New York Academy of Sciences* 1387(1): 44–53.
- Sittig DF and Singh H (2020) COVID-19 and the need for a national health information technology infrastructure. *JAMA* 323(23): 2373–2374. DOI: [10.1001/jama.2020.7239](https://doi.org/10.1001/jama.2020.7239)
- Tomines A, Readhead H, Readhead A, et al. (2013) Applications of electronic health information in public health: uses, opportunities & barriers. *EGEMS* 1(2): 1019–1019.
- Ulahannan JP, Narayanan N, Thalath N, et al. (2020) A citizen science initiative for open data and visualization of COVID-19 outbreak in Kerala, India. *Journal of the American Medical Informatics Association* 27(12): 1913–1920.
- Vogel J, Brown JS, Land T, et al. (2014) MDPHnet: secure, distributed sharing of electronic health record data for public health surveillance, evaluation, and planning. *American Journal of Public Health* 104(12): 2265–2270.
- Wager KA, Wickham Lee F and Glaser JP (2013) *Health Care Information Systems: A Practical Approach for Health Care Management*. United States of America, New Jersey: Jossey-Bass.
- Willis SJ, Cocoros NM, Randall LM, et al. (2019) Electronic health record use in public health infectious disease surveillance, USA, 2018-2019. *Current Infectious Disease Reports* 21(10): 32. DOI: [10.1007/s11908-019-0694-5](https://doi.org/10.1007/s11908-019-0694-5)
- Wissel BD, Van Camp PJ, Kouril M, et al. (2020) An interactive online dashboard for tracking COVID-19 in U.S. counties, cities, and states in real time. *Journal of the American Medical Informatics Association* 27(7): 1121–1125. DOI: [10.1093/jamia/ocaa071](https://doi.org/10.1093/jamia/ocaa071)

- World Health Organization (2022) *WHO Coronavirus (COVID-19) dashboard*. Geneva, Switzerland: WHO. Available at: <https://covid19.who.int/>(accessed 20 April 2022).
- World Health Organization (2020) *Coronavirus disease (COVID-19) outbreak situation*. Available at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (accessed 08 February 2022).
- Yahyavi A, Hemmati N, Derakhshan P, et al. (2021) Angiotensin enzyme inhibitors and angiotensin receptor blockers as protective factors in COVID-19 mortality: a retrospective cohort study. *Internal and Emergency Medicine* 16: 883–893. Doi: [10.1007/s11739-020-02523-9](https://doi.org/10.1007/s11739-020-02523-9)
- Zarei J, Dastoorpoor M, Jamshidnezhad A, et al. (2021) Regional COVID-19 registry in Khuzestan, Iran: a study protocol and lessons learned from a pilot implementation. *Informatics in Medicine Unlocked* 23: 100520.
- Zhu N, Zhang D, Wang W, et al. (2020) A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine* 382: 727–733. DOI: [10.1056/NEJMoa2001017](https://doi.org/10.1056/NEJMoa2001017)