Managing Pandemics with Health Informatics: Successes and Challenges

Mujeeb A. Basit1,2, Christoph U. Lehmann1,3, Richard J. Medford1,2
1 Clinical Informatics Center, UT Southwestern, Medical Center, Dallas, TX, USA
2 Department of Internal Medicine, UT Southwestern, Medical Center, Dallas, TX, USA
3 Departments of Pediatrics, Population & Data Sciences, and Bioinformatics, UT Southwestern, Medical Center, Dallas, TX, USA

Summary

Introduction: The novel COVID-19 pandemic struck the world unprepared. This keynote outlines challenges and successes using data to inform providers, government officials, hospitals, and patients in a pandemic.

Methods: The authors outline the data required to manage a novel pandemic including their potential uses by governments, public health organizations, and individuals.

Results: An extensive discussion on data quality and on obstacles to collecting data is followed by examples of successes in clinical care, contact tracing, and forecasting. Generic local forecast model development is reviewed followed by ethical consideration around pandemic data. We leave the reader with thoughts on the next inevitable outbreak and lessons learned from the COVID-19 pandemic.

Conclusion: COVID-19 must be a lesson for the future to direct us to better planning and preparing to manage the next pandemic with health informatics.

Keywords
COVID-19, SARS-COV-2, 2019 novel coronavirus disease, pandemics, mass immunization, COVID-19 testing, public reporting of healthcare data, communicable diseases, medical informatics

1 Introduction
The world was unprepared for the 2019-2021 Coronavirus Disease (COVID-19) pandemic caused by the novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) when the disease was first identified in Wuhan, China in December 2019. At the time this manuscript was completed, more than 68 million people worldwide have been infected and global deaths exceed 1.55 million [1].

As of December 8 2020, COVID-19 still has no proven cure but the first vaccine doses were administered in the United Kingdom [2]. The absence of a reliable treatment highlights the importance of public health principles such as rapid case identification, monitoring, contact tracing, and quarantining/isolating to stop the spread of a communicable disease as well as non-pharmaceutical interventions (NPI) such as mask wearing, social distancing, and hand washing. The most important tool for governments and public health officials in the fight against a pandemic are data to aid in planning, decision-making, and measuring effectiveness. We discuss the data needed, the failures and challenges to obtain them, the successes associated with the use of data, and the ethical considerations around pandemic data.

2 What Data are Needed in a Pandemic and for What Purposes?
When exploring data required to managing a pandemic, the different consumers of the data must be considered. Table 1 outlines potential consumers of data (National, State, Regional/Local agencies, Hospital/Healthcare enterprise, and patients/individuals). For each data element, we outline the possible uses of the data by the consuming entity. We included data elements that historically have been used during pandemics such as new infections, testing data, ventilators, personal protective equipment (PPE), health care resources, and patients [3]. For the COVID-19 pandemic, novel data not available in the past such as mobility data measured by cell phones, and social media and crowdfunding data were also included. Mobility data allow an estimation of movement, exposure to risky locations (e.g., restaurants and gyms), or participation in larger gatherings (e.g., protests or Canadian or US Thanksgiving holidays) [4]. Social media data can be used to assess a population’s attitude towards non-pharmaceutical interventions like social distancing [5] and sentiment towards the pandemic [6]. Crowdfunding can indicate a community in distress and can be used to identify communities that require additional services to handle the pandemic [7].

3 Failures or Challenges to Obtain Accurate Data
Critical to responding and combating the SARS-CoV-2 pandemic is the ability to gather and aggregate accurate and timely test results and to identify patients quickly and correctly to allow for rapid isolation and tracing of contacts. While seemingly simple, this task
requires clear and disambiguated data requests, common data definitions, and standardized variables. Further, the task requires trained staff that can label and reliably handle laboratory specimens, aggregate data correctly, create and transmit reports, and identify process problems or inaccurate information.

### 3.1 Lack of Strategy and Data Harmonization

In countries like the US, where COVID-19 public health strategies lack centralization and harmonization, efforts to combat the virus largely failed and led to some of the highest number of COVID-19 infections in the world [8]. In the beginning of the pandemic, there was widespread disagreement about which date should be attributed to a COVID-19 test (e.g., onset of symptoms, date the specimen was obtained, date the test was run, date the results were reported), with dates often stretching over a range of weeks making interpretation challenging.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Data use by data consuming entities.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>National</strong></td>
<td><strong>State</strong></td>
</tr>
<tr>
<td>New Infections / new positive tests</td>
<td>Disease prediction Tracing Resource utilization forecasting National PPE stockpile forecasting</td>
</tr>
<tr>
<td>Tests Performed</td>
<td>Test Use Prediction Test supply allocation Disease Prediction</td>
</tr>
<tr>
<td>PPE availability/burn rate</td>
<td>National PPE stockpile forecasting</td>
</tr>
<tr>
<td>Hospital Beds (Available/Occupied) / Currently Admitted Patients</td>
<td>National resource allocation (e.g., military health care providers)</td>
</tr>
<tr>
<td>Previously Admitted or COVID positive Patients</td>
<td>Herd immunity Vaccination target rates</td>
</tr>
<tr>
<td>Immunizations</td>
<td>Herd immunity Vaccination target rates Immunization priority</td>
</tr>
<tr>
<td>Mobility</td>
<td>Disease prediction</td>
</tr>
<tr>
<td>Non-Pharmaceutical Interventions (mask mandate, social distancing, mandatory closures, …etc.)</td>
<td>Disease prediction Resource utilization forecasting</td>
</tr>
<tr>
<td>Social Media</td>
<td>Disease prediction Resource utilization forecasting Public health messaging Public sentiment measurement</td>
</tr>
<tr>
<td>Crowdfunding</td>
<td>Marker of community in distress</td>
</tr>
</tbody>
</table>

IMIA Yearbook of Medical Informatics 2021
3.2 Lack of a Skilled Workforce

Requesting COVID-19 data requires domain knowledge in infectious diseases as well as informatics. As anticipated, lack of understanding of available data and processes resulted in requests at various levels of the government (federal, state, and local levels) that were frequently ambiguous and not congruent - adding to the data collection effort at the frontlines, which required a substantial workforce during the pandemic [11]. The results of a lack of rigor and preciseness in data requests were best represented by a data request to hospitals made by the head of the White House Corona Task force [12]. This request did not consider the condition that an academic medical center or a laboratory may also function as a reference laboratory which resulted in duplicate reporting of tests, difficult or impossible to adjudicate once data were reported.

3.3 Failing Information Systems

Entities that require COVID-19 data reporting frequently request a variety of formats over various time periods and with varying deadlines, which can add significantly to the reporting burden for healthcare organizations. Add to this the fact that the automation, the integration, and the interoperability of electronic systems available to public health departments, at local and state levels, substantially vary as public health information technology infrastructures and their capabilities diverge widely. Further, health department infrastructures (at least in the US) tend to be outdated and fail frequently. In Dallas county (the home of the authors), the health department limited access to its information systems during the daytime to reduce user load in an otherwise unreliable system.

3.4 Reporting Errors, Delays, and Misrepresentations

In the US, COVID-related deaths most likely reached 100,000 three weeks before the official milestone, because of reporting delays and failures to accurately attribute all COVID-related deaths to the virus [13, 14]. The Centers for Disease Control and Prevention (CDC) conflated results of polymerase chain reaction (PCR) tests, which test for the presence of the virus, with antibody tests, which test for past infections, resulting in an inaccurate timeline of infections and in an overinflating of testing capabilities [15-17]. Texas, Virginia, and Vermont were accused of misrepresenting their data to inflate the size of their testing programs [18]. The US State of Georgia displayed COVID-19 data in a graph that suggested that case numbers were dropping when, in fact, they were not [19]. In the absence of universal definitions, there are numerous ways to calculate positivity rates netting different results that are used [20]. Misrepresentations - whether intentionally or accidentally - further hindered the ability to interpret data correctly.

More troubling, however, are delays in data reporting. When the pandemic surges, reporting delays lead to underestimation of the actual disease prevalence, which hinders forecasting the spread or resurgence of the disease. Missing positive test results may make an analysis conclude that the disease is spreading at a different rate than it actually is. Calculating the effective reproduction number (Rt) based on data with missing results may suggest a transmission rate that would quench the infection while using all test results may lead to the conclusion that the disease is actually spreading and result in incorrect policy decisions [11].

4 Data Successes

Despite the challenges associated with the COVID-19 pandemic, many institutions, hospital systems, states, and countries have risen to the challenge by leveraging informatics to facilitate clinical care, ensure appropriate testing, perform contact tracing, develop forecasting tools to support decision makers, and advance prediction through artificial intelligence techniques.

4.1 Clinical Care

Driven by the need to minimize SARS-CoV-2 spread by enacting NPIs like physical distancing, the adoption of technologies such as telehealth platforms worldwide marked a significant departure from traditional clinical care [21]. Whereas telehealth had traditionally poor penetration in most countries largely due to regulatory and reimbursement hurdles [22, 23], the rapid development of technological infrastructures and training providers on best practices signified a transformational shift in the delivery of care. The UC Health System, an integrated health system spanning over three states, transitioned within a two-week period from 350 to over 20,000 weekly telehealth visits [24]. Similarly, the Veterans Health Administration (VA) moved to virtual outpatient visits (defined as video or telephone care), from 14% to 58% over a three-month period (out of 5,400,878 outpatient visits) [25]. Providers were forced to learn new skill-sets including developing rapport over video calls, reinventing the physical exam, and engaging learners. Furthermore, clinicians asked for remote patient monitoring of blood pressure, pulse oximetry, heart rate, and blood glucose in order to better serve their patients, using infrastructures that likely existed at many institutions, but were underutilized [26].

4.2 Contact Tracing

Contact tracing, an essential tool in managing an outbreak, is a labor-intensive process that involves identification and information retrieval of individuals, who have potentially come into contact with an infected person. Given the velocity and volume of new COVID-19 infections, hospitals, public health systems, and countries were forced to adapt and develop new methods of digital contact tracing. Apple Inc. and Google LLC partnered to develop a Bluetooth smartphone solution with privacy at the forefront of their solution, garnering
adoption from countries like Germany, Italy, and the United Kingdom [27]. South Korea leveraged its experience from the Middle East Respiratory Syndrome (MERS) and enhanced its contact tracing methods to include card transactions, closed circuit television, global positioning systems, and medical records to identify and verify each individual’s movement [28]. The National Centre for Infectious Diseases in Singapore - Singapore’s national COVID-19 screening referral center - compared a wearable real-time locating system (RTLS) tag and contact-based tracing app to traditional electronic health record (EHR) methods and demonstrated increased sensitivity of the RTLS-based system in identifying provider-contact events within their facility, acknowledging the difficulty to employ similar technology in a community setting [29]. Taiwan used its national immigration and customs and health insurance databases to identify individuals at risk due to recent travel to “hot-spots” and quarantined these individuals, ensuring quarantine compliance through tracking of mobile devices (see the section “6. Ethical Challenges” below) [30].

4.3 Forecasting
Using traditional and novel sources of data, e.g., smartphone mobility data, COVID-19-related search trends, social media data, and app-based vital sign and symptom data streams, modeling the viral dynamics and transmission have helped inform policymakers at local and national levels to successfully enact public health measures and understand sources of “super-spreader” events [31, 32]. Models have been used to inform policy decisions in Australia, Canada, and China [33-35]. Similarly, modeling techniques were used to identify African countries at moderate to high risk of COVID-19 importation into their respective countries, which highlighted the need for increased preparedness and infrastructure to address the impending risk [36]. On a local level, models have successfully been employed to aid hospital systems in understanding staffing, surge and physical capacity, and supply chain demands [37].

4.4 Scientific Discovery through Artificial Intelligence
Besides forecasting, the COVID-19 pandemic has highlighted the use of artificial intelligence (AI) to aid in clinical care, from prediction of a positive SARS-CoV-2 test to deep learning systems for diagnosis and prognostication [38,39]. Moreover, AI’s potential to help in reduction of financial, material, and human labor costs in low to middle income countries where the necessary infrastructure, resources, and workforce may not be given has also been highlighted [40]. One particular area where AI has been extremely beneficial is therapeutics and drug repurposing (i.e., identifying effective agents from existing approved or investigational drugs). For example, Remdesivir, a drug initially used for Ebola virus, has been repurposed for COVID-19 and applications of AI through target and network analysis (text mining and proteomics), compound-based virtual screening (ligand-based similarity and molecular docking), and drug-specific filters (drug-drug interactions, synergy, interactions and genomics) have enabled clinical trials with drugs including dexamethasone and Baricitinib [41-44].

5 Creating Local Prediction Models
Data is essential for monitoring the spread of COVID-19 and anticipating when changes to processes, capacity management, and policies are needed to curb the growth. Unfortunately, during this crisis, the efficient, accurate, and timely testing and dissemination of results have been a struggle (see the section “3. Failures or Challenges to obtain Accurate Data”). Therefore, local and regional informatics and healthcare communities must join forces to share data, analytics, and forecasts to increase community awareness, inform hospital administrations, and guide policy leaders on the appropriate steps needed to avoid stressed healthcare systems and clinicians to prevent poor outcomes such as increased morbidity [45].

For a model to be useful for local and regional prediction models, it must contain data from various local and regional sources, including cell phone mobility data, business and school closure data, COVID-19-like illness symptoms data, clinic visit data, community and hospital-based testing data, emergency department and hospitalization visit data, intensive care unit admission data, discharge data, and death information. Transforming the data into a type-2-dimensional model captures the data change over time, allowing the calculation of the number of results reported late and the modeling of the data pipeline reliability. Tools like a susceptible-exposed-infectious-removed (SEIR) model calculate transition functions allowing prognostication of hospitalizations, ICU admissions, and recovery and death. COVID-19 forecasting reports (Figure 1 shows a sample) allow diverse audiences such as news media, community leaders, and lay people to understand the pandemic’s status and its trajectory.

Hospital level data focus on patient characteristics, testing data, and encounter level data. These data elements are important to modeling, easy to collect, and there is generally low resistance to sharing. Encounter data allow the calculation of transition probabilities from the emergency department to the hospital to the intensive care unit to death or convalescence at the individual encounter level while controlling for patient characteristics and risk factors. The length of stay information at the encounter level allows tracking of capacity needs and changes over the course of an outbreak (example in Figure 2).

6 Ethical Challenges
6.1 Rationing Resources
At the time we are writing this paper in early December 2020, the US hospitals are quickly reaching capacity and ICU beds are becoming scarce [46] or unavailable. The looming threat of limited resources and potential deliberate withholding of care for patients considered less likely to survive requires attention. As the Working
Group on Ethical, Legal, and Social Issues at the American Medical Informatics Association suggested, scoring tools already existing in EHRs, such as the Acute Physiology and Chronic Health Evaluation (APACHE), Sequential Organ Failure Assessment (SOFA), quick SOFA (qSOFA), modified SOFA (mSOFA), Pediatric Risk of Mortality (PRISM), and others were developed for outcome prediction and were not intended to serve as tools in resource allocation or triage of patients in the context of deciding on scarce resource allocations in a novel pandemic. This includes decisions, on who will receive an ICU bed, be intubated, or resuscitated [47]. In case of rationing resources, we refer the readers to the framework developed by Douglas White at the University of Pittsburgh for rationing ventilators and critical care beds [48], and to the policy for fair allocation of medications [49].

![Fig. 1 Dallas County lab confirmed COVID-19 occupied beds and ICU occupied beds. The prediction and uncertainty bands are created using a stochastic Markov SEIR model.](image1)

![Fig. 2 Length of stay boxplot of COVID-19 admissions with color gradients showing weekly admission numbers. This visualization was made possible through the collaboration of regional clinical partners and aggregation of the encounter level data model that contains one entry per hospitalization.](image2)
6.2 Advanced Directives
The pandemic, where many patients died alone in hospitals because friends and family were banned from visiting, has made abundantly clear that a push is needed to actively include advanced directives into the EHR to aid in end-of-life decision-making, especially when surrogates are not available to be questioned about the patient’s final wishes. The use of advance directives and code status in the EHR remains low. Ground work to include advanced directives has been laid by members of the American Medical Informatics Association’s Ethics Committee [50].

6.3 Testing and Tracing
The WHO promoted isolation, testing, and tracing as a response to the COVID-19 pandemic. The premise of this approach is to test individuals, isolate those who tested positive, and trace their contacts [51]. Apple and Google released COVID-19 tracking software whereby users, who opt in to install the software, are notified by the app if a person, who has been in close contact, discloses to the app that he/she tested positive for COVID-19 [52]. A Kaiser Family Foundation survey found that Americans were split with 47% unwilling to download the app rendering its utility low. Concerns about government overreach, distrust of large private technology companies, and worries that such an application could be used to monitor quarantine [53] will continue to keep the utilization of this app low. While voluntary surveillance tools are of concern, even more concerning are hidden or covert tools to collect data which may be abused by private interests or autocratic governments.

The collection and use of data through phones, closed circuit cameras, banking apps and credit cards, and purchases in the battle against the pandemic can easily become very intrusive, abusive, and put individuals at risk especially when the data are sold or utilized for unintended purposes. Data may put individuals at risk for discrimination, bias, denial of employment or insurance, ostracizing, and financial disadvantages. Using data in the fight of the pandemic requires an understanding of consumer protection and how much public health needs may infringe on autonomy and the right to privacy. However, these arguments do not prevent the creation of a transparent and secure contact-tracing app with such technologies as block chain and personal health records (https://www.nature.com/articles/d41586-020-03518-4).

6.4 Telehealth’s Unintended Consequences
We adapted to the COVID-19 pandemic by transitioning care to telehealth. However, this change has introduced unintended consequences. First, there is the phenomenon of ‘Zoom Fatigue’, which has been described as “the tiredness, anxiety, or worry resulting from overusing these virtual platforms” [54]. With over 300 Million daily participants in Zoom meetings [55] and many other available platforms such as Teams, BlueJeans, GoToMeeting to name a few, the problem is widespread. One explanation for the drain on humans is audio delay. A one second audio/video delay results in increased frustration [56] and 1.2 seconds result in the communication partners to be perceived as less attentive, conscientious, and extroverted [57]. Functional MRI data showed that viewing a video is associated with less activation of reward centers in the brain than face-to-face interactions [58]. In addition, telehealth delivery could be associated with increased provider burnout due to other factors such as screen fatigue, decreased attention due to the sedentary position (physical activity is associated with a 40% lower risk of fatigue [59]), increased miscommunication, and the difficulty to emote or read emotions through this medium.

7 Discussion
7.1 Thoughts for the Next Pandemic
Unlike prior pandemics, EHR transaction systems were already in place to process the new testing requirements for the COVID-19 pandemic. However, these EHR systems did not receive and implement new ontology concepts [60] or recipients/consumers of information until later in the pandemic [61]. Even when the endpoints for the delivery of knowledge and information were established, they were limited by inconsistent reporting standards, changing requirements, multiple competing reporting agencies, and poor definitions. Informatics working with local, state, and national reporting agencies improved workflows and data definition to better monitor the pandemic at increasingly granular levels resulting in better informed policy makers and improved policies [62]. There are many informatics lessons to be learned from this pandemic and fortunately it appears (with vaccines being rolled out) that we will have the opportunity—and the moral obligation—to improve the core health reporting systems and response infrastructures to better detect and respond to the next outbreak, which is inevitable.

Before the next outbreak, we must achieve a significant improvement to our reporting policies and technology systems. We need processes and rules for rapid development of ontology definitions, widely available testing algorithms, test reporting for positive and negative results, encounter level reporting related to the disease, symptom reporting for all patients, and consistent death reporting. Such frameworks should be in place for common infectious diseases such as influenza, and will be easy to scale when new infectious agents emerge. The reporting should be transparent with key metrics including time to result report, successful contact tracing, distribution of exposure types, and percent closed loop reporting. Each institution that sends results to a central reporting agency should get those same results returned by the local contact tracing office to validate that results were received correctly and to determine time to report. As discussed earlier, with COVID-19, there have been many barriers and delays to reporting with a significant portion of results not reported in a timely manner or not reported at all - defeating any contact tracing efforts. With early vaccinations, we are starting to see a similar problem with provided immunizations failing to be reported. The informatics community has already achieved much to improve the reporting, modeling, and educating during this pandemic but there
is more to be done to create a reporting standard with appropriate transparency, checks, and automations to reduce the impact of the next outbreak.

7.2 Learning from the COVID-19 Pandemic

Worldwide, there have been stark differences in the response to the COVID-19 pandemic resulting in widely different outcomes. Based on measures like cases or deaths per million people, the United States, the United Kingdom, and Sweden performed worse than other industrial nations like Canada, Japan, Germany, and South Korea. Once vaccines have been administered to most of the world’s population, countries will likely commence a period of re-assessing their pandemic responses and will identify failures with likely consequences for responsible institutions and individuals, whose actions are identified as insufficient, negligent, inapt, or even criminal. Comparing national outcomes will play an important role in identifying successful actors and those leaders, who failed their populaces.

After the Great Depression in the US, Franklin D. Roosevelt introduced a series of sweeping policies, programs, and reforms that addressed social inequities, workers’ rights, racism, and discrimination. If past is prologue, we can anticipate reforms across the globe as a result of this pandemic. The 1918 Spanish Flu pandemic shed light on the impact of different local public health responses and the deaths resulting from mistakes made by local officials. The 1957 Asian Flu pandemic strengthened the role of the World Health Organization and emphasized the beneficial nature of vaccination programs [63]. The COVID-19 pandemic is likely to draw similar conclusions; however, unlike in the past respective analysis will also shine a light on inequities, injustices, biases, and social determinants of health. Exploring the reasons why certain groups carried a disproportionately larger burden of morbidity, mortality, and suffering will be addressed and will result in societal changes and transformations in the delivery of health care.

In the spirit of learning from the past, it is imperative that informaticians across the world continue their role of data stewards and historians to this pandemic and carry on collecting, recording, and analyzing data with the intent of gathering new knowledge and translating it into a better, faster, and more successful response to the next pandemic, which is inevitable.

8 Conclusion

In a pandemic, data are the best tool available to clinicians, patients, administrators, and government officials. As COVID-19 demonstrated, the world was woefully under-prepared despite years of predictions [64]. Despite the under preparedness, clinicians and informaticians came together to develop ways to collect, standardize, and disseminate data needed to make necessary decisions. COVID-19 must be a lesson for the future to direct us to better planning and preparing to manage the next pandemic with health informatics.

References

8. Dreher A. State has been over-reporting negative COVID-19 results The error came from the inclusion of negative antibody test results with negative COVID-19 results The error came from the inclusion of negative antibody test results with negative COVID-19 results. HeraldNet Everett, Washington [cited 2020 Jul 2]. https://www.heraldnet.com/northwest/state-has-been-over-reporting-negative-covid-19-results/
ing-picture-of-pandemic


Managing Pandemics with Health Informatics: Successes and Challenges


Correspondence to:
Mujeeb A. Basit, MD, MMSc
5323 Harry Hines Blvd 8578
Dallas, TX 75390, USA
Tel: +1 214 648 1303
E-mail: Mujeeb.Basit@UTSouthwestern.edu