Editorial

Research and Data Mining During the COVID-19 Pandemic

Hale Z. Toklu, PhD, MPharm,[1,2] Rakesh Prashad, MD, FACP, FACC, FSCAI[3]

Author affiliations are listed at the end of this article.

Correspondence to:
Hale Z. Toklu, PhD, MPharm
Director of Research,
HCA North Florida Division
Assoc. Prof., University of Central Florida College of Medicine, Dept. of Clinical Sciences
(hale.toklu@hcahealthcare.com)

Abstract

Description

With the recent advances in information and technology systems, most developed countries have invested in building advanced systems for the management of electronic medical records. If the infrastructure of these systems are well designed, they both serve as an information resource for routine patient care and also serve as the primary backbone for medical research. This evidence became clear during the recent coronavirus pandemic, which has been a worldwide challenge since the beginning of 2020. Rapid spread of SARS–CoV-2 infections all over the world has resulted in tremendous health, economic and social ramifications, including social distancing, travel restrictions and closing of schools and businesses. Medical educational activities have shifted towards telemedicine, online learning and web-based meetings and conferences to prevent virus spread. While experimental lab research slowed, research with electronic medical records and databases accelerated in order to investigate the risk factors associated with COVID-19 and clinical management strategies to combat the disease. The urgency for COVID-related research has also lead to inconsistent and sometimes inaccurate output.

Data mining efforts aim to formulate, analyze and implement basic induction processes that facilitate the extraction of meaningful information and knowledge from unstructured data. Data mining extracts patterns, changes, associations and anomalies from large data sets. With the recent advances in information and technology systems, most developed countries invested in building advanced systems for the management of electronic medical records. When the infrastructure of these systems are well-designed, they can serve as an information resource for routine patient care and medical research.[1]

The importance of a well-built infrastructure of electronic medical records (EMR) has become evident during the coronavirus pandemic, which has been a worldwide challenge since the beginning of 2020. As of October, COVID-19 has infected 35.3 million people worldwide with over one million deaths. The rapid spread of SARS-CoV-2 infections all over the world has had severe health, economic and social consequences, including social distancing, travel restrictions and the closing of schools and businesses. Medical educational activities have shifted towards telemedicine, online learning and web-based meetings and conferences to prevent virus spread.

While experimental lab research has slowed, research with electronic medical records and databases has accelerated to investigate the factors associated with COVID-19 and its clinical management, leading to strategies to combat the disease. In many academic institutions, new projects that require an in-person presence are not permitted, except for COVID-19-related basic and clinical research, which is permitted and encouraged. "Dry" laboratory research is carried out remotely unless specific dispensa-

tion is warranted, and active therapeutic clinical trials remain ongoing. Research activities that are deemed critical are permitted with some restrictions and regulations.[2]

On the other hand, database research, especially via electronic medical records, has increased significantly. Today's technology for collecting and analyzing data has shown great promise for using innovative, data-driven tools to fight the coronavirus. However, the urgency for quick turnover of COVID-19-related studies has affected the quality of studies. As of September 2020, 33 papers have been retracted; 3 have been temporarily retracted and 2 are under investigation due to quality concerns.[3] Even top journals with high impact factors have retracted articles.

Inconsistent results from these studies have undermined trust in the pharmaceutical industry, research organizations and scientific publications. Indeed, the inconsistency and conflicting data can be explained . First of all, the diversity of populations for retrospective cohorts is an important factor, e.g., the age of patients in China is comparably younger than the ones in the UK. This affects different outcomes in terms of morbidity and mortality. Another reason is the availability of different diagnostic tests and treatment regimens across countries. When polypharmacy is added to the scene, it becomes even more challenging to evaluate the actual efficacy of drugs as well as the side effects. The drug-induced side effects overlapped with the disease symptoms in some cases, e.g., sodium imbalance in COVID-19. Is the imbalance due to sepsis or hydration? Should we suppress or not suppress the immune system? Shall we give steroids or hydroxychloroquine in combination with remdesevir or other antivirals? Achieving optimal medication was complex under normal circumstances in patients who require polypharmacy. Optimal medication has become even more challenging during the pandemic given the diverse population with multiple co-morbidities and chronic conditions.[4]

Another problem regarding research during the pandemic is the urgency in research. The desire to publish multiple studies rapidly combined with insufficient peer review undermines the quality of research during the pandemic. This theory also applies to the poor design and administrative review of research protocols. The urgency to publish what is available in the database may result in researchers missing key pieces of the puzzle.

Despite all the problems working with retrospective data extracted from electronic medical records, it is still valuable. Given the fact that we knew little about the SARS-Co-V infections before the pandemic, it is noteworthy how much data has accumulated in such a short period of time, which has been a monumental help in developing treatment strategies to fight against the virus.

Conflicts of Interest

The authors declare they have no conflicts of interest.

Author Affiliation

1. Department of Clinical Sciences, University of Central Florida College of Medicine, Orlando, FL
2. HCA Healthcare Graduate Medical Education
3. Internal Medicine Residency Program, University of Central Florida/HCA Healthcare Ocala Regional Medical Center, Ocala, FL

References

1. Judd RM, Kim RJ. Electronic Medical Records and Medical Research Databases - Can They Be Synonymous? *US Cardiology*. 2006;2(1):136-139. https://www.uscjournal.com/journals/editions/us-cardiology-2005

2. Omary MB, Eswaraka J, Kimball SD, Moghe PV, Panettieri RA Jr, Scotto KW. The COVID-19 pandemic and research shutdown: staying safe and productive. *J Clin Invest*. 2020;130(6):2745-2748. https://doi.org/10.1172/jci138646

3. Retracted coronavirus (COVID-19) papers. Retraction Watch. https://retractionwatch.com/retracted-coronavirus-covid-19-papers/. Accessed October 5, 2020.

4. Ailabouni NJ, Hilmer SN, Kalisch L, Braund R, Reeve E. COVID-19 Pandemic: Considerations for Safe Medication Use in Older Adults with Multimorbidity and Polypharmacy [published online ahead of print, 2020 Apr 30]. *J Gerontol A Biol Sci Med Sci*. 2020;glaa104. https://doi.org/10.1093/gerona/glaa104