THE VALUE OF A REGIONAL 'LIVING' COVID-19 REGISTRY AND THE CHALLENGES OF KEEPING IT ALIVE

Posted on August 2, 2021 by Matthew

Category: Summer 2021

By John Hanna, MD[*]; Tara Chen, MD[*]; Carlos Portales-Castillo, MD[*]; Mina Said, MD; Rene Bulnes, MD; Donna Newhart, MS; Lucas Sienk, MSBA; Katherine Schantz, MPA; Kathleen Rozzi, CCS-P; Karan Alag, MD; Jonathan Bress, MD; and Emil Lesho DO

*Equal contributors

Abstract

Background: The availability of accurate, reliable, and timely clinical data is crucial for clinicians, researchers, and policymakers so that they can respond effectively to emerging public health threats. This was typified by the recent SARS-CoV-2 pandemic and the critical knowledge and data gaps associated with novel Coronavirus 2019 disease (COVID-19).

We sought to create an adaptive, living data mart containing detailed clinical, epidemiologic, and outcome data from COVID-19 patients in our healthcare system. If successful, the approach could then be used for any future outbreak or disease.

Methods: From 3/13/2020 onward, demographics, comorbidities, outpatient medications, along with 75 laboratory, 2 imaging, 19 therapeutic, and 4 outcome-related parameters, were manually extracted from the electronic medical record (EMR) of SARS-CoV-2 positive patients. These parameters were entered on a registry featuring calculation, graphing tools, pivot tables, and a macro programming language. Initially, two internal medicine residents populated the database, then professional data abstractors populated the registry. Clinical parameters were developed with input from infectious diseases and critical care physicians and using a modified COVID-19 worksheet from the U.S. Centers for Disease Control and Prevention (CDC). Registry contents were migrated to a browser-based, metadata-driven electronic data capture software platform. Eventually, we developed queries and used various business intelligence (BI) tools which enabled us to semi-automate data ingestion of 147 clinical and outcome parameters from the EMR, via a large U.S. hospital-based, service-level, all-payer database. Statistics were performed in R and Minitab.

Results: From March 13, 2020 to May 17, 2021, 549,691 SARS-CoV-2 test results on 236,144 distinct patients, along with location, admission status, and other epidemiologic details are stored on the cloud-based BI platform. From March 2020 until May 2021, extraction of clinical-epidemiologic parameter had to be performed manually. Of those, 543 have had >/=75 parameters fully entered in the registry. Ten clinical characteristics were significantly associated with the need for hospital admission. Only one characteristic was associated with a need for ICU admission. Use of supplemental oxygen, vasopressors and outpatient statin were associated with increased mortality.

Initially, 0.5hrs -1.5 hours per patient chart (approximately 450-575 person hours) were required to manually extract the parameters and populate the registry. As of May 17, 2021, semi-automated data ingestion from the U.S. hospital all-payer database, employing user-defined queries, was implemented. That process can ingest and populate the registry with 147 clinical, epidemiologic, and

outcome parameters at a rate of 2 hours per 100 patient charts.

Conclusion:

A living COVID-19 registry represents a mechanism to facilitate optimal sharing of data between providers, consumers, health information networks, and health plans through technology-enabled, secure-access electronic health information. Our approach also involves a diversity of new roles in the field, such as using residents, staff, and the quality department, in addition to professional data extractors and the health informatics team.

Initially, due to the overwhelming number of infections that continues to accelerate, and the labor/time intense nature of the project, only a small fraction of all patients with COVID-19 had all parameters entered in the registry. Therefore, this report also offers lessons learned and discusses sustainability issues, should others wish to establish a registry. It also highlights the registry's local and broader public health significance. Beginning in June 2021, whole-genome sequencing results such as lineages harboring important viral mutations, or variants of concern will be linked to the clinical meta-data.

Keywords: SARS-CoV-2, COVID-19, registry, mortality, epidemiology, genomics, genomic epidemiology, electronic medical record

Background

One dilemma in the attempt to deliver state-of-the-art therapeutics or understand emerging novel infectious pathogens is that the most current data is not always rapidly accessible to clinicians as soon as it becomes available. Hence, living systematic reviews have arisen as a potential solution to narrow the gap between evidence and practice.[1, 2]

Along the same lines, having live access with continuous updates of a patient data base using a standardized, well designed format and a user friendly, interactive software which allows us to also to more rapidly hypothesis test or identify patterns relevant to prevention motivated this effort. Additionally, data from one country or U.S. locality may not be generalizable to others, and data and experience from non-university or rural settings may be underrepresented in the medial literature, or not covered by media.[3] Furthermore, rural areas are currently experiencing some of the biggest increases in new SARS-CoV-2 infections. Fourth, on June 19, 2020, the National Academies of Sciences, Engineering, and Medicine hosted a public meeting on "Data Needs to Monitor the Evolution of SARS-CoV-2". Presenters agreed that regional surveillance nodes were needed. Last, the recently described "tragic data gap", and the federal curtailment of reporting COVID-19 data to the national Health Safety Network, also provided motivation for this intervention.[4] In the current climate, archiving detailed patient data in retrievable, easily analyzable, and share-friendly formats has become crucial for informing responses to current and future pandemics.[4]

Recent guidance[5] and events[6, 7, 8] also underscore the added value having private, nongovernmental alternatives for collecting and analyzing large scale epidemiologic data.

We sought to implement an adaptive, 'living' registry capable of capturing detailed epidemiologic and clinical information from every patient diagnosed with SARS-CoV-2 infection, similar to and more granular than the Danish COVID-19 Cohort or TriNetX network. Examples of the living approach include the living rapid reviews in the Annals of Internal Medicine, and the living systematic review of the University of Toronto regarding secondary infections in patients with COVID-19. However, these are literature reviews and not registries.

Our ultimate goal was to have a minable source of detailed information updated on an ongoing basis, with minimal human effort, and linkable to SARS-CoV-2 whole genome sequences, and other registries across the United States. Perhaps, becoming one of many extra-governmental surveillance nodes in a regional or national network. Additionally, if successful, the approach could then be used for any future outbreak or disease regardless of pathogen or etiologic agent.

Setting

The healthcare system consists of five acute care hospitals (ACH) and six long- term care facilities (LCTF) totaling 1056 ACH and 800 LCTF beds. It spans eleven suburban, rural, and urban counties in the Finger-Lakes Region of NY.

Methods

All patients in the above healthcare system who had a positive test for SARS-CoV-2 from 03/13/2020 onward were slated for inclusion in the RRH-COVID-19 Registry. Every patient who undergoes SARS-CoV-2 testing in the health system is captured in a cloud-based business intelligence (BI) platform that permits self-service data visualization and guided analytics (QlikView, Lund, Sweden). From this database, the medical record number of each patient is used to drive further queries (either manually or semi-automated) of the enterprise electronic medical record (EMR) (Epic, Verona, WI).

In mid-March 2020, after the first case in our community, demographics, comorbidities, outpatient medications, along with 75 laboratory, 2 imaging, 19 therapeutic, and 4 outcome-related parameters were manually extracted from the electronic medical record of SARS-CoV-2 positive patients. These parameters were developed based on input from infectious diseases and critical care specialists and manually entered on a registry featuring calculation, graphing tools, pivot tables, and a macro programming language (Excel). Initially, two internal medicine residents populated the database, then professional data abstractors populated the registry.

When the U.S. Centers for Disease Control and Prevention (CDC) released their COVID-19

Surveillance Worksheet [9], we used that as a guide to ensure the parameters we were collecting

were aligned with national reporting efforts. To those parameters, we added several more such as laboratory values, imaging results, location of admission, length of stay and other outcome data. The initial platform (Excel) was abandoned and the contents transferred to a browser-based, metadata-driven electronic data capture software platform (REDCap, Vanderbilt University, TN)

Eventually, we developed queries and used various business intelligence tools which enabled us to semi-automate data ingestion of 147 clinical and outcome parameters (except imaging data) from the EMR via a large U.S. hospital-based, service-level, all payer database (Premier, Charlotte, NC).[10]

Statistics were performed in R and Minitab.

Results

From March 13, 2020 to May 17, 2021, the healthcare system performed 549,691 SARS-CoV-2 tests on 236,144 distinct patients. 30,213 of those tested positive. Everyone, including the negatives are stored, along with location, admission status, and other epidemiologic details are stored on the cloud-based BI platform. Users can partially customize dashboards to view trends, geography, laboratory details and census data. From March 2020 until May 2021, extraction of clinical-epidemiologic parameter had to be performed manually. Of those, 543 have had the more than 75 parameters fully entered in the registry. It took from 0.5hrs -1.5 hours per chart (approximately 450-575 person hours to extract the parameters and enter them in the registry.

As of May 17, 2021, semi-automated data ingestion from the U.S. hospital all-payer database, employing user-defined queries, was implemented. Using that process, all 147 clinical, epidemiologic, and outcome parameters can be accomplished at a rate of 2 hours per 100 patients.

Descriptive statistics of the first cohort of patients are presented in **Table 1**, with mortality associations in **Table 2**. The average follow-up period for those was 25 days (range 21-34 days). Ten characteristics were significantly associated with the need for hospital admiss

ion: age, male gender, occupation as a healthcare worker, diabetes, hypertension, cardiovascular disease, kidney disease, cancer, use of statins, use of ACEI-ARBs, and acid suppressant use. The only characteristic associated with need for ICU admission was a history of close contact w

ith a SARS-CoV-2 infected person. The use of supplemental oxygen, vasopressors, and outpatient statin was associated with increased mortality (**Table 2**).

Discussion

The registry revealed local patterns not apparent from less granular databased such as the CDC or county/state health departments, or from reports of other populations. For example, the average length of stay (LOS) for all admitted patients was 8 days (SD 8.6). The average LOS for patients who did and did not require ICU-level care was 13.5 (SD 10.4) and 5.9 (SD 6.8), respectively. This is significant because remdesivir was not available during the follow up period for this initial cohort,

and the average LOS without it was already shorter than the 15-day outcome measure used in the final report.[11] Another example is a meta-analysis stating that ACEI or ARB use was associated with a lower risk for severe illness. According to the authors, those results "do not provide enough evidence to draw conclusions about the potential efficacy of these medications in treating COVID-19".[2] In contrast, we found use of angiotensin inhibitors and angiotensin receptor blockers was higher in admitted vs. ED treated-and-released patients (p ≤ 0.001). A third example involves a Chinese report which showed that statin treatment was associated with lower mortality.[12] However in our population, statin use was higher in admitted patients compared to patients who did not require admission (all p ≤ 0.001). Statin use was also higher in those who died than those who survived (**Table 2**) Last, in New York City, viral load was correlated with risk of intubation, but in our more rural and suburban area, we observed that viral load did not correlate with severity of illness.[13, 14]

Despite its benefits, designing and implementing a registry or specific data mart can be fraught- especially for hospitals with limited health informatics and financial support. Registry challenges, pitfalls, and threats to sustainability are presented in **Table 3**. Manual data extraction into the original spreadsheet became prohibitively labor intensive and analytically unmanageable as the number of new cases rapidly increased. Real-time abstraction needed to be suspended for several weeks while contents were transferred to the metadata-driven electronic data capture software platform. Initially, limited by a dependence on manual data extraction registry personnel were able to manually populate all the parameters mentioned above from only a small fraction of COVID-19 patients. Several reasons account for this: the labor intense nature, the unprecedented numbers of patients with the target condition due to pandemic surges or waves, and the number of clinical epidemiologic and outcome parameters we tried to capture. Designers and developers have to keep this in mind as they balance the desire to capture more patents with less data, or less patients with more parameters.

For diseases with lower incidence rates, the sustainability of the initial approach would have been achievable, but the approach became unsustainable given epidemic/pandemic-level volumes of new cases daily and weekly. For optimal sustainability, two full-time experienced data extractors, automation and machine-based learning would be required. The semi-automated process we used also has limitations. Using that, we still have to rely on manual chart review for obtaining imaging data, home/outpatient medication use, and certain health habits such as smoking and alcohol consumption.

Conclusion

A living COVID-19 registry represents a mechanism to facilitate optimal sharing of data between providers, consumers, health information networks, and health plans through technology-enabled, secure-access electronic health information. Our approach also involves a diversity of new roles in

the field such as using residents, staff, and the quality department, in addition to professional data extractors and the health informatics team.

However, due to the overwhelming number of infections that accelerated over the course of the pandemic, and the labor/time intense nature of the project, only a fraction of all patients with COVID-19 had all parameters entered in the registry. Therefore, this report also offers lessons learned and discusses sustainability issues, should others wish to establish a registry. It also highlights the registry's local and broader public health significance.

Robust registries from large EMR vendors take time to build and release, at times waiting for maturing knowledge makes them unavailable to initial nodes of outbreak. At such places, regional nodal registries are critical in early understanding and management of the pandemic. The progress may be limited initially due to the manual process, but they critically guide development of robust semi-automated or automated solutions. In addition, the results from early studies provide direction to broader preparedness and deployment programs. During the early part of a pandemic, focus changes to operational readiness and management, requiring lots of IT infrastructure bandwidth. At such times, having a manual process can be crucial to success of registry effort.

Going forward, the registry is well suited to address a range of hypothesis and is being leveraged by other researchers in our system. It provides a resource for researchers, policy makers, and surge planning. Once established, the informatics processes and parameters can be applied to patients infected with any high-consequence or novel pathogen. Soon, the clinical-demographic metadata will be linked to whole genome sequencing data for patients that had their virus sequenced. A growing number (currently 650) of patients have had whole genome sequence performed on their samples. This approach has been successfully used to improve outcomes for drug-resistant bacteria[15], and our next goal is to do the same for SARS-CoV-2.

Author Biographies

John Hanna, MD, is an internal medicine resident in the medicine department at Rochester Regional Health in Rochester, New York.

Tara Chen, MD, *is an internal medicine resident in the Medicine Department at Rochester Regional Health in Rochester, New York.*

Carlos Portales-Castillo, MD, *is an internal medicine resident in the Medicine Department at Rochester Regional Health in Rochester, New York.*

Mina Said, MD, *is an internal medicine resident in the Medicine Department at Rochester Regional Health in Rochester, New York.*

Rene Bulnes, MD, *is an internal medicine resident in the Medicine Department at Rochester Regional Health in Rochester, New York.*

Donna Newhart, MS, is a performance improvement manager at the Quality and Safety Institute at Rochester Regional Health, Rochester, New York.

Lucas Sienk, MSBA, is a performance improvement coordinator at the Quality and Safety Institute at Rochester Regional Health, Rochester, New York.

Katherine Schantz, MPA, is a senior director in the Quality Reporting Department at Rochester Regional Health in Rochester, New York.

Kathleen Rozzi, CCS-P, is a senior director in the Quality Reporting Department at Rochester Regional Health in Rochester, New York.

Karan Alag, MD, is the medical director of health informatics in the Medicine Department at Rochester Regional Health in Rochester, New York.

Jonathan Bress, MD, is a nephrologist in the Medicine Department at Rochester Regional Health in Rochester, New York.

Emil Lesho, DO, is a healthcare epidemiologist in the Medicine Department at Rochester Regional Health in Rochester, New York.

References

1. Elliott JH, Turner T, Clavisi O, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. PLoS Med. 2014;11:e1001603.
2. Mackey K, King VJ, Gurley S, et al. Risks and impact of angiotensin-converting enzyme inhibitors or angiotensin-receptor blockers on SARS-CoV-2 infection in adults: a living systematic review. Ann Intern Med. 2020;173:195-203.
3. McMinn S, Stone W, Westwood R. As coronavirus cases surge, NPR examines hospital capacity. 2020. https://www.npr.org/2020/07/28/896088067/as-coronavirus-cases-surge-npr-examines-hospital-capacity. Accessed 28 October 2020.
4. Schneider EC. Failing the test - the tragic data gap undermining the U.S. pandemic response. N Engl J Med. 2020;383:299-302.
5. Health and Human Services. COVID-19 guidance for hospital reporting and FAQs for hospitals, hospital laboratory, and acute care facility data reporting updated July 29, 2020. 2020. https://www.hhs.gov/sites/default/files/covid-19-faqs-hospitals-hospital-laboratory-acute-care-facility-data-reporting.pdf. Accessed 28 October 2020.
6. Wamsley L. Fired Florida data scientist launches a coronavirus dashboard of her own. 2020. https://www.npr.org/2020/06/14/876584284/fired-florida-data-scientist-launches-a-coronavirus-dashboard-of-her-own. Accessed 28 October 2020.
7. Huang P, Simmons-Duffin S. White House strips CDC of data collection role for COVID-19 hospitalizations. 2020.

https://www.npr.org/sections/health-shots/2020/07/15/891351706/white-house-strips-cdc-of-data-collection-role-for-covid-19-hospitalizations. Accessed 28 October 2020.

8. Arvisais-Anhalt S, Lehmann CU, Park JY, et al. What the coronavirus disease 2019 (COVID-19) pandemic has reinforced: the need for accurate data. Clin Infect Dis. 2021;72:920-923.

9. Centers for Disease Control and Prevention. Human Infection with Coronavirus Disease 2019 (COVID-19) Surveillance Worksheet https://www.cdc.gov/coronavirus/2019-ncov/downloads/php/COVID19-Worksheet-CSV-annotated-20201Jan15.pdf. Accessed 18 May 2021.

10. Premier. Premier healthcare database (COVID-19): data that informs and performs. https://learn.premierinc.com/white-papers/premier-healthcaredatabase. Accessed 18 May 2021.

11. Beigel JH, Tomashek KM, Dodd LE, et al. Remdesivir for the treatment of Covid-19 - final report N Engl J Med. 2020; doi:10.1056/NEJMoa2007764.

12. Zhang XJ, Qin JJ, Cheng X, et al. In-hospital use of statins is associated with a reduced risk of mortality among individuals with COVID-19. Cell Metab. 2020;32:176-187.

13. Lesho E, Reno L, Newhart D, et al. Temporal, spatial, and epidemiologic relationships of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) gene cycle thresholds: a pragmatic ambi-directional. Clin Infect Dis. 2020; doi:10.1093/cid/ciaa1248.

14. Magleby R, Westblade LF, Trzebucki A, et al. Impact of SARS-CoV-2 viral load on risk of intubation and mortality among hospitalized patients with Coronavirus Disease 2019. Clin Infect Dis. 2020; doi:10.1093/cid/ciaa851.

15. Lesho EP, Waterman PE, Chukwuma U, et al. The antimicrobial resistance monitoring and research (ARMoR) program: the U.S. Department of Defense response to escalating antimicrobial resistance. Clin Infect Dis. 2014;59:390-7.

There are no comments yet.