Minnesota Electronic Health Record Consortium COVID-19 Project: Informing Pandemic Response Through Statewide Collaboration Using Observational Data

Tyler N. A. Winkelman, MD, MSc[1,2] (iD); Karen L. Margolis, MD[3];
Stephen Waring, DVM, PhD[4]; Peter J. Bodurtha, MPP[2]; Rohan Khazanchi, BA[2,5,6];
Stefan Gildemeister, MA[7]; Pamela J. Mink, PhD, MPH[7]; Malini DeSilva, MD, MPH[3];
Anne M. Murray, MD[8,9]; Nayanjot Rai, MPH[10]; Julie Sonier, MPA[11];
Claire Neely, MD[12]; Steven G. Johnson, PhD[13]; Alanna M. Chamberlain, PhD, MPH[14];
Yue Yu, PhD[14]; Lynn M. McFarling, MD[15]; R. Adams Dudley, MD, MBA[13,16,17];
and Paul E. Drawz, MD, MHS, MS[10], for the Minnesota EHR Consortium

Abstract

Objective: Robust disease and syndromic surveillance tools are underdeveloped in the United States, as evidenced by limitations and heterogeneity in sociodemographic data collection throughout the COVID-19 pandemic. To monitor the COVID-19 pandemic in Minnesota, we developed a federated data network in March 2020 using electronic health record (EHR) data from 8 multispecialty health systems.

Materials and Methods: In this serial cross-sectional study, we examined patients of all ages who received a COVID-19 polymerase chain reaction test, had symptoms of a viral illness, or received an influenza test from January 3, 2016, through November 7, 2020. We evaluated COVID-19 testing rates among patients with symptoms of viral illness and percentage positivity among all patients tested, in aggregate and by zip code. We stratified results by patient and area-level characteristics.

Results: Cumulative COVID-19 positivity rates were similar for people aged 12-64 years (range, 15.1%-17.6%) but lower for adults aged ≥65 years (range, 9.3%-10.7%). We found notable racial and ethnic disparities in positivity rates early in the pandemic, whereas COVID-19 positivity was similarly elevated across most racial and ethnic groups by the end of 2020. Positivity rates remained substantially higher among Hispanic patients compared with other racial and ethnic groups throughout the study period. We found similar trends across area-level income and rurality, with disparities early in the pandemic converging over time.

Practice Implications: We rapidly developed a distributed data network across Minnesota to monitor the COVID-19 pandemic. Our findings highlight the utility of using EHR data to monitor the current pandemic as well as future public health priorities. Building partnerships with public health agencies can help ensure data streams are flexible and tailored to meet the changing needs of decision makers.

Keywords
COVID-19, infectious diseases, health disparities, public health surveillance, health informatics

In the United States, COVID-19 incidence and mortality has illuminated inequities by race and ethnicity, age, and socioeconomic status, driven by long-standing social inequality and disparities in access to and quality of care.[1,2] The United States has among the highest per-capita rates of SARS-CoV-2 infection and death in the world, despite spending the most on health care per capita.[3] Effective and rapid disease and syndromic surveillance, supported in many countries through investments in public health infrastructure, is critical to manage public health crises such as the current pandemic.[4,5] However, disease and syndromic surveillance methods in the United States that include laboratory information, comprehensive sociodemographic data, and clinical information have been inadequate during the pandemic, in part because of chronic underfunding.[6-8] Data on incident cases, diagnostic testing, and health care use by race and ethnicity, language, and other important subgroups are critical to identify outbreaks, locate areas with inadequate testing, allocate resources, and guide nonpharmaceutical public health interventions.

Electronic health records (EHRs) are an emerging tool for public health surveillance.[9-11] EHR data include demographic information (including variables often unavailable through traditional public health surveillance methods, such as race, ethnicity, and language), zip codes and geolocations, health care use across encounter types, diagnosis codes, procedure codes, and laboratory orders and results. However, the breadth of EHR data is not readily available to most public health agencies in the United States.[11] When available, they are rarely customizable at the pace required during a pandemic.

In March 2020, health care professionals, researchers, and leaders from Minnesota health systems, statewide health care organizations with expertise in multisystem collaboration and measurement (ie, Institute for Clinical Systems Improvement, MN Community Measurement), and the Minnesota Department of Health (MDH) met with the goal of developing a Minnesota EHR Consortium (hereinafter, "the Consortium") to study the epidemiology of chronic conditions such as cardiovascular disease, hypertension, and substance use disorders. The COVID-19 pandemic provided an unanticipated motivation to expand the focus of this collaborative and rapidly establish a statewide EHR-based COVID-19 surveillance system in partnership with MDH.

The primary goal of the Consortium's COVID-19 Project was to address gaps in traditional public health surveillance and inform a coordinated statewide response to the pandemic. A key principle was protecting the security and privacy of individual data by using a federated data network. The Consortium's early objectives were (1) to develop syndromic surveillance definitions to detect viral symptoms and determine testing rates among people with viral symptoms, (2) to contribute a more equitable and complete understanding of disease variation by demographic characteristics and clinical comorbidities than what was available through existing data infrastructure, and (3) to demonstrate the ability to work and govern collaboratively and flexibly to meet the needs of health system and public health leaders.

The Consortium worked with representatives from MDH to identify how EHR data could augment ongoing surveillance efforts. In March/April 2020, MDH had access to all COVID-19 laboratory results; billing data for admission, discharge, and transfer files for emergency department visits and hospitalizations; and information from patient interviews. The Consortium was able to augment this information with data on patients with COVID-19–like symptoms in outpatient and telehealth settings, including patients who had not received COVID-19 testing. The Consortium was also able to provide data on demographic, comorbidity, and area-level characteristics among tested patients with viral symptoms and patients with confirmed COVID-19 on a week-by-week basis; these data were not systematically or completely collected by MDH.

We describe the development process of the Consortium's COVID-19 Project, provide initial results, and discuss future implications of our work, including how health system and public health collaborations can augment existing surveillance efforts and inform short- and long-term responses to COVID-19 and other pressing public health challenges.

Materials and Methods

Study Design and Setting

Across all 8 health systems involved in the Consortium at the time of this study, we obtained deidentified EHR data for patients of all ages who received a COVID-19 polymerase chain reaction (PCR) test, had symptoms of a viral illness (hereinafter, viral symptoms), or received an influenza test from January 3, 2016, through November 7, 2020. We updated

[1] Division of General Internal Medicine, Department of Medicine, Hennepin Healthcare, Minneapolis, MN, USA

[2] Health, Homelessness, and Criminal Justice Lab, Hennepin Healthcare Research Institute, Minneapolis, MN, USA

[3] HealthPartners Institute, Minneapolis, MN, USA

[4] Essentia Health, Essentia Institute of Health, Duluth, MN, USA

[5] School of Public Health, University of Minnesota, Minneapolis, MN, USA

[6] College of Medicine, University of Nebraska Medical Center, Omaha, NE, USA

[7] Minnesota Department of Health, Saint Paul, MN, USA

[8] Division of Geriatrics, Department of Internal Medicine, Hennepin Healthcare, Minneapolis, MN, USA

[9] Berman Center for Outcomes and Clinical Research, Hennepin Healthcare Research Institute, Minneapolis, MN, USA

[10] Division of Nephrology and Hypertension, University of Minnesota Medical School, Minneapolis, MN, USA

[11] MN Community Measurement, Minneapolis, MN, USA

[12] Institute for Clinical Systems Improvement, Minneapolis, MN, USA

[13] Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA

[14] Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA

[15] CentraCare, St. Cloud, MN, USA

[16] Department of Medicine, University of Minnesota Medical School, Minneapolis, MN, USA

[17] Minneapolis Veterans Affairs Medical Center, Minneapolis, MN, USA

Corresponding Author:

Tyler N. A. Winkelman, MD, MSc, Hennepin Healthcare, Department of Medicine, Division of General Internal Medicine, 701 Park Ave, S2.309, Minneapolis, MN 55415, USA.

Email: tyler.winkelman@hcmed.org

estimates weekly beginning in March 2020. Institutional review boards at each participating site reviewed the project. All applications were approved, deemed to be exempt nonhuman subjects research, or deemed to be exempt public health surveillance.

Demographic, Clinical, and Area-Level Socioeconomic Characteristics

Patient demographic and clinical characteristics including age (all ages), sex (male/female), race and ethnicity (White, Black, American Indian/Alaska Native, Asian/Pacific Islander, multiracial, Hispanic, other/unknown/missing), language (Spanish, Somali, English, other), need for an interpreter (yes/no), comorbidities (asthma, chronic obstructive pulmonary disease, HIV, cancer, heart disease, diabetes, and chronic kidney disease), and patient zip code were identified from EHR data. Approximately 15% of patients did not have data on language; missingness for the remaining demographic variables was generally ≤5%.

Race and ethnicity definitions were mutually exclusive and reflected self-reported identities collected from the EHR. Hispanic patients were all labeled as Hispanic, whereas non-Hispanic patients were designated according to race. The multiracial category reflects non-Hispanic patients with multiple race designations. *International Classification of Diseases, Tenth Revision* (ICD-10) codes were used to define comorbidities.[12,13] Neighborhood-level socioeconomic status was defined at the zip code level using median annual household income from the 2015-2019 American Community Survey (ACS).[14] Rurality was defined at the zip code tabulation area (ZCTA) level using 2010 rural–urban commuting area (RUCA) codes from the US Department of Agriculture's Economic Research Service[15] and percentage of patients living in rural areas from the 2015-2019 ACS.[14] ZCTAs were labeled as urban if they were classified as urban by RUCA and had <50% of the population living in a rural area, rural if they were classified as rural by RUCA and had ≥50% of the population living in a rural area, exurban if they were classified as urban by RUCA and had ≥50% of the population living in a rural area, and small town if they were classified as rural by RUCA and had <50% of the population living in a rural area.

COVID-19 and Viral Illness Case Definitions

The Consortium defined COVID-19–positive patients as those who received a positive test result for SARS-CoV-2 using the PCR test. The definition of viral illness was updated and retroactively applied to previous time points as reports of symptom prevalence among various populations with confirmed COVID-19 were published.[16-20] We defined viral symptoms as the presence of any of the following ICD-10 codes: fever (R50.9, R50.81), cough (R05), respiratory failure (J96.00, J96.01, J96.02), influenza-like illness (J09-J11),

viral pneumonia (J12, J16-J18), or coronavirus (U07.1, B97.21, B97.29, B34.2, J12.81).[13] Additional data from laboratory orders identified whether patients were symptomatic or asymptomatic at the time of SARS-CoV-2 testing.

Outcomes

Outcomes reported by the Consortium on a weekly basis included COVID-19 testing rates among patients with viral symptoms (ie, COVID-19 tests among patients with viral symptoms/total patients with viral symptoms) and percentage test positivity in aggregate and at the zip code level (ie, COVID-19–positive patients/total patients tested for COVID-19; referred to as "positivity rates" hereinafter). We also reported COVID-19 tests and positivity rates among symptomatic and asymptomatic patients, year-over-year counts of patients with viral symptoms, and similar data for influenza in our online dashboard; however, we did not include these data in the results of this evaluation. Symptoms at the time of testing were identified if a health care provider selected "symptomatic" in a PCR order. We reported results by race and ethnicity, language, age, sex, comorbidities, and area-level characteristics such as income and rurality.

Data Aggregation and Analysis

We used a federated data network, and our methods matured from March through November 2020. We initially developed a detailed codebook that documented inclusion and exclusion criteria and data formats. Each week, Consortium sites produced a summary file containing counts of patients with confirmed COVID-19 stratified by demographic and area-level characteristics, patients with viral symptoms, and the number of COVID-19 tests and corresponding positivity rates (overall and among those with viral symptoms). Sites shared summary data using standardized templates; summary data were centrally merged and analyzed for dissemination.

During the first 6 months of the project, the effort of adding new variables to the data model and ensuring data integrity across all sites grew in complexity. Because sites did not share a common data model, we developed centrally managed code to deploy a novel standardized, patient-level data model in November 2020 that functioned as a project-specific common data model. Variable creation was based on definitions outlined in a codebook developed by the project leaders. Given the evolving nature of the pandemic, sites updated the laboratory definition for a COVID-19 PCR test. For each site, abstracted patient data were included in the intermediate file if the patient had a COVID-19 test, met the viral symptoms case definition, or had a documented influenza test. The intermediate file was organized at the patient-week level; thus, a single patient could contribute multiple rows of data in the intermediate file if the patient had encounters or laboratory studies that met our inclusion criteria in

multiple weeks. The standardized code linked each site's intermediate file to an area-level file with ZCTA information, produced the summary tables, and transferred summary data for central merging and analysis. Procedures were developed to identify potential reporting errors including spot checks, week-over-week comparisons, visual inspection of site-specific trends, and comparisons with internal and statewide estimates.

We performed statistical analyses in R version 4.0.3 (R Foundation for Statistical Computing). We created maps using Power BI (Microsoft Corp) to demonstrate rates of COVID-19 test positivity, COVID-19 case and testing rates, and viral illness rates by ZCTA. Summary results were initially shared as documents but later uploaded each week to an online dashboard and discussed in collaboration with health system and MDH representatives. We used *Morbidity and Mortality Weekly Report* (*MMWR*) weeks for reporting time. An *MMWR* week is the week of the epidemiologic year used by the Centers for Disease Control and Prevention for disease reporting.

Results

As of December 2020, eight large MN Health Systems (Allina Health, Children's Minnesota, Essentia Health, HealthPartners, Hennepin Healthcare, M Health Fairview, CentraCare, and Mayo Clinic) had contributed data to the Consortium's COVID-19 Project. Approximately 50% of all COVID-19 tests with a positive result in Minnesota were included in Consortium data, with greatest representation from the Twin Cities metropolitan area.

Demographic, Clinical, and Area-Level Characteristics

We examined demographic, clinical, and area-level characteristics for patients who received a COVID-19 test, COVID-19–positive patients with corresponding positivity rates, patients with viral symptoms, and patients with viral symptoms who received a COVID-19 test with corresponding testing rates for *MMWR* weeks 42-45 (October/November 2020; Table). During this period, 13.6% of all COVID-19 test results were positive and 60.4% of patients with viral symptoms were tested. The percentage of positive test results was similar among all people aged 12-64 years who were tested (range, 15.1%-17.6%) but was lower among children aged <12 years (range, 6.2%-9.6%) and adults aged ≥65 years (9.3%-10.6%). We found notable racial and ethnic disparities in positivity rates early in the pandemic that closed toward the end of 2020 because of rising positivity rates across most racial and ethnic groups, but positivity rates remained substantially higher among Hispanic patients compared with patients in other racial and ethnic groups (Figure 1A; Supplemental Tables S1 and S2). Disparities in positivity rates by language persisted throughout 2020 (Table; Supplemental Tables S1 and S2). The

patient population that received COVID-19 testing was similar to the overall care-seeking population at Consortium sites (Supplemental Table S3).

COVID-19 Test Positivity by ZCTA-Level Characteristics

Disparities by area-level income and rurality also converged during the study period (Figure 1B, C). This convergence occurred because of rising levels of COVID-19 positivity across all income quartiles and geographic regions. Rising levels of positivity, despite high levels of testing (testing levels during *MMWR* weeks 42-45 [Table] were higher than during *MMWR* weeks 13-16 [Supplemental Table S1]), indicate increased case growth. In April/May 2020 (*MMWR* weeks 14-22), patients from neighborhoods in the lowest quartile of median annual household income had positivity rates as high as 22.9%, whereas patients from neighborhoods in the highest income quartile had positivity rates as high as 9.1%. By November 2020, these disparities were attenuated in the wake of increasing positivity rates across all neighborhoods regardless of median annual household income. We observed a similar pattern by rurality. From April to June 2020 (*MMWR* weeks 18-23), urban areas and small towns were disproportionately affected. Urban areas experienced a second wave in the summer (*MMWR* weeks 28-36), with positivity rates rising from <5.0% in June (*MMWR* weeks 24-25) to 6.9% in August (*MMWR* weeks 31-32). Positivity rates gradually increased in exurban and rural areas from June (*MMWR* weeks 23-26) through September (*MMWR* weeks 36-39). As with median annual household income, positivity rates increased steeply in all areas by November (*MMWR* weeks 44-48), regardless of rurality.

COVID-19 Positivity and Testing by ZCTA

We found geographic disparities in positivity and testing rates among patients with viral symptoms (Figure 2). In *MMWR* week 47, we found areas of high positivity in northern Minnesota and in the outer suburbs of the Minneapolis–St Paul metropolitan area. Similarly, the percentage of viral cases tested varied by zip code. In the Minneapolis–St Paul metropolitan area, most neighborhoods had a testing rate >50%, with some >75%. However, the percentage of viral cases tested was <40% in some neighborhoods.

Discussion

Our findings highlight the shifting nature of disparities throughout the pandemic, including persistently elevated COVID-19 positivity rates among Hispanic patients. We were able to identify target populations in which either case growth was elevated or levels of access to testing were low, which informed outreach efforts by health care systems and public health officials to address key drivers of elevated

Table. Demographic, clinical, and area-level characteristics associated with COVID-19 testing and positivity among the overall study population[a] and among patients meeting viral case criteria, as of *MMWR* weeks 42-45,[b] Minnesota, 2020

Variable	No. (%) of COVID-19 tests received	No. (%) of positive COVID-19 test results	COVID-19 test positivity rate, %	No. (%) of patients with symptoms of a viral illness[c]	No. (%) of patients with symptoms of a viral illness who were tested	Testing rate among patients with symptoms of a viral illness, %
Total	280 456 (100.0)	38 250 (100.0)	13.6	78 100 (100.0)	47 206 (100.0)	60.4
Age, y						
0-4	12 388 (4.4)	765 (2.0)	6.2	5661 (7.2)	3651 (7.7)	64.5
5-11	15 124 (5.4)	1448 (3.8)	9.6	3718 (4.8)	2644 (5.6)	71.1
12-18	18 205 (6.5)	2786 (7.3)	15.3	3983 (5.1)	2988 (6.3)	75.0
19-24	21 835 (7.8)	3835 (10.0)	17.6	5451 (7.0)	3791 (8.0)	69.5
25-44	83 030 (29.6)	12 561 (32.8)	15.1	20 368 (26.1)	13 713 (29.0)	67.3
45-64	73 594 (26.2)	11 219 (29.3)	15.2	21 173 (27.1)	12 147 (25.7)	57.4
65-74	29 635 (10.6)	3156 (8.3)	10.6	9135 (11.7)	4398 (9.3)	48.1
≥75	26 648 (9.5)	2482 (6.5)	9.3	8618 (11.0)	3875 (8.2)	45.0
Race and ethnicity[d]						
White	223 156 (79.6)	29 373 (76.8)	13.2	59 540 (76.2)	35 558 (75.3)	59.7
Black	16 136 (5.8)	2183 (5.7)	13.5	5959 (7.6)	3577 (7.6)	60.0
American Indian/Alaska Native	2069 (0.7)	210 (0.5)	10.1	620 (0.8)	395 (0.8)	63.7
Asian/Pacific Islander	6983 (2.5)	992 (2.6)	14.2	2364 (3.0)	1309 (2.8)	55.4
Multiracial	2752 (1.0)	285 (0.7)	10.4	1150 (1.5)	562 (1.2)	48.9
Hispanic	12 300 (4.4)	2799 (7.3)	22.8	4792 (6.1)	3209 (6.8)	67.0
Other/unknown/missing	17 027 (6.1)	2396 (6.3)	14.1	3668 (4.7)	2596 (5.5)	70.8
Sex						
Male	120 252 (42.9)	18 212 (47.6)	15.1	35 564 (45.5)	21 800 (46.2)	61.3
Female	160 120 (57.1)	20 038 (52.4)	12.5	42 531 (54.5)	25 401 (53.8)	59.7
Language						
Spanish	4142 (1.5)	1317 (3.4)	31.8	2109 (2.7)	1422 (3.0)	67.4
Somali	1998 (0.7)	374 (1.0)	18.7	950 (1.2)	510 (1.1)	53.7
English	263 792 (94.1)	35 098 (91.8)	13.3	72 521 (92.9)	43 730 (92.6)	60.3
Other[e]	2767 (1.0)	456 (1.2)	16.5	1284 (1.6)	631 (1.3)	49.1
No. of comorbidities[e]						
0	195 594 (69.7)	29 371 (76.8)	15.0	48 237 (61.8)	32 401 (68.6)	67.2
1	39 051 (13.9)	4601 (12.0)	11.8	12 935 (16.6)	6865 (14.5)	53.1
≥2	45 851 (16.3)	4287 (11.2)	9.3	16 957 (21.7)	7963 (16.9)	47.0
Rurality[f]						
Rural	37 436 (13.3)	6093 (15.9)	16.3	8810 (11.3)	6136 (13.0)	69.6
Small town	33 616 (12.0)	4693 (12.3)	14.0	7636 (9.8)	5329 (11.3)	69.8
Exurban	16 477 (5.9)	2599 (6.8)	15.8	4023 (5.2)	2570 (5.4)	63.9
Urban	171 297 (61.1)	21 815 (57.0)	12.7	50 454 (64.6)	30 552 (64.7)	60.6
Socioeconomic status (median annual household income)[g]						
Quartile 1 (low)	46 187 (16.5)	5937 (15.5)	12.9	13 084 (16.8)	8083 (17.1)	61.8
Quartile 2	38 818 (13.8)	5031 (13.2)	13.0	11 268 (14.4)	6815 (14.4)	60.5
Quartile 3	44 639 (15.9)	6065 (15.9)	13.6	12 421 (15.9)	7586 (16.1)	61.1
Quartile 4 (high)	57 882 (20.6)	7360 (19.2)	12.7	17 633 (22.6)	10 601 (22.5)	60.1

[a]Data source: Deidentified data from 8 health systems on patients who received a COVID-19 polymerase chain reaction test, had symptoms of a viral illness, or received an influenza test from January 3, 2016, through November 7, 2020.

[b]*Morbidity and Mortality Weekly Report* (*MMWR*) week is the week of epidemiological year used by the Centers for Disease Control and Prevention for disease reporting.

[c]The viral case definition included the presence of any of the following *International Classification of Diseases, Tenth Revision* (ICD-10) codes: fever (R50.9, R50.81), cough (R05), respiratory failure (J96.00, J96.01, J96.02), influenza-like illness (J09-J11), viral pneumonia (J12, J16-J18), or coronavirus (U07.1, B97.21, B97.29, B34.2, J12.81).[13]

[d]Definitions of race and ethnicity are mutually exclusive and reflect self-reported identities collected from the electronic health record. Hispanic patients, regardless of race, were categorized as Hispanic; non-Hispanic patients were categorized according to race. The multiracial category includes non-Hispanic patients with multiple race designations. The other/unknown/missing category includes those who belong to any race or ethnicity other than the mentioned categories, did not know their race or ethnicity, or did not provide the answer.

[e]Comorbidities were defined using ICD-10 codes. Comorbidities included asthma, chronic obstructive pulmonary disease, HIV, cancer, heart disease, diabetes, and chronic kidney disease.

[f]Rurality was defined at the zip code tabulation area (ZCTA) level using 2010 rural–urban commuting area (RUCA) codes from the US Department of Agriculture's Economic Research Service[15] and percentage of patients living in rural areas from the American Community Survey (ACS).[14] ZCTAs were labeled as urban if they were classified as urban by RUCA and had <50% of the population living in a rural area, rural if they were classified as rural by RUCA and had ≥50% of the population living in a rural area, exurban if they were classified as urban by RUCA and had ≥50% of the population living in a rural area, and small town if they were classified as rural by RUCA and had <50% of the population living in a rural area.

[g]Neighborhood-level socioeconomic status was defined at the ZCTA level using median annual household income from the 2015-2019 ACS.[14]
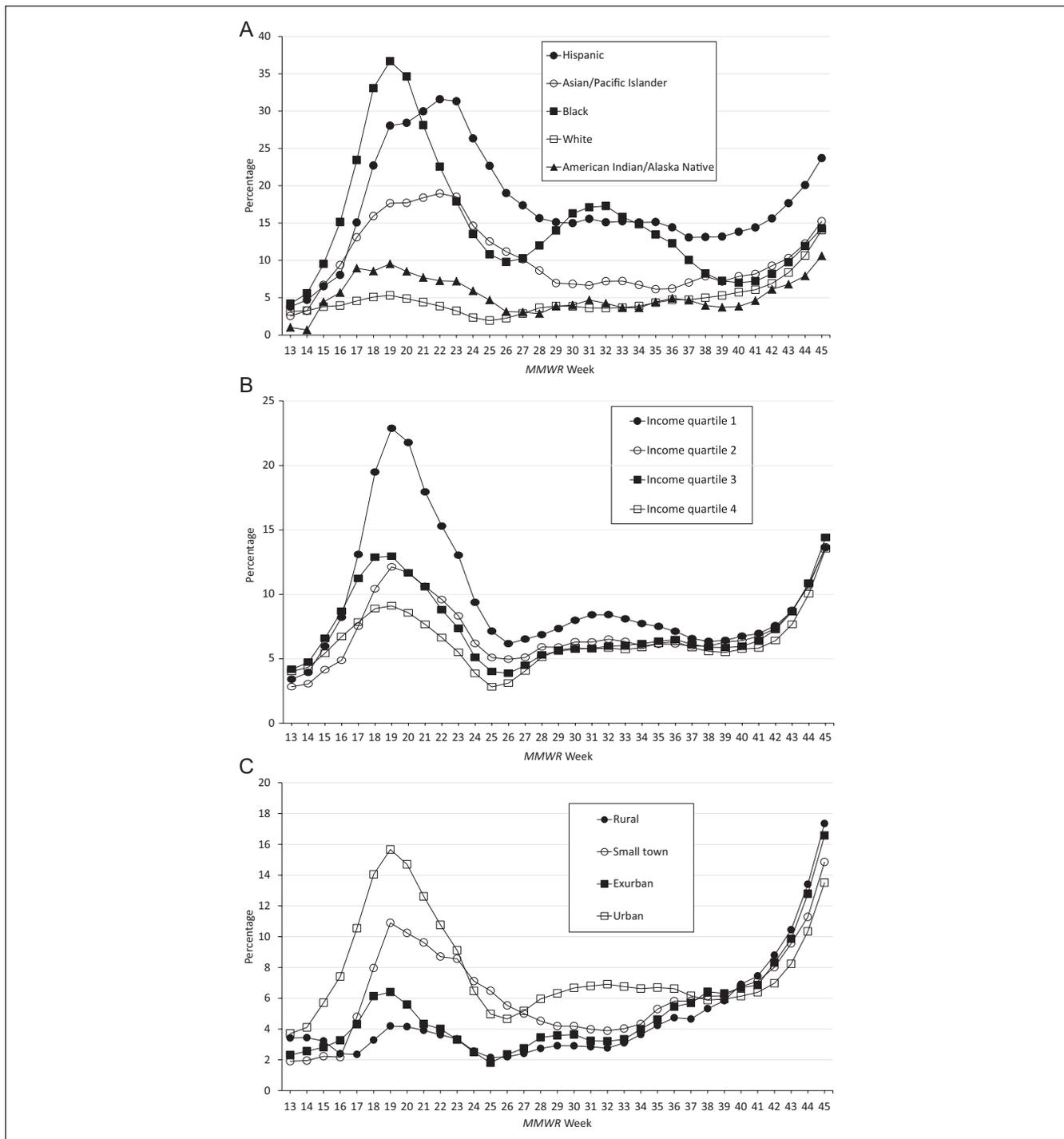
Figure 1. COVID-19 positivity rates within the Minnesota Electronic Health Record (EHR) Consortium, by (A) individual race and ethnicity, (B) zip code tabulation area (ZCTA)–level socioeconomic status, and (C) ZCTA-level rurality, Minnesota, 2020. Estimates are 3-week rolling averages. Definitions of race and ethnicity are mutually exclusive and reflect self-reported identities collected from the EHR. Hispanic patients, regardless of race, were categorized as Hispanic; non-Hispanic patients were categorized according to race. Neighborhood-level socioeconomic status was defined at the ZCTA level using median annual household income from the 2015-2019 American Community Survey (ACS).[14] Rurality was defined at the ZCTA level using 2010 rural–urban commuting area (RUCA) codes from the US Department of Agriculture's Economic Research Service[15] and percentage of patients living in rural areas from the 2015-2019 ACS.[14] ZCTAs were labeled as urban if they were classified as urban by RUCA and had <50% of the population living in a rural area, rural if they were classified as rural by RUCA and had ≥50% of the population living in a rural area, exurban if they were classified as urban by RUCA and had ≥50% of the population living in a rural area, and small town if they were classified as rural by RUCA and had <50% of the population living in a rural area. An *MMWR* (*Morbidity and Mortality Weekly Report*) week is the week of the epidemiologic year used by the Centers for Disease Control and Prevention for disease reporting.
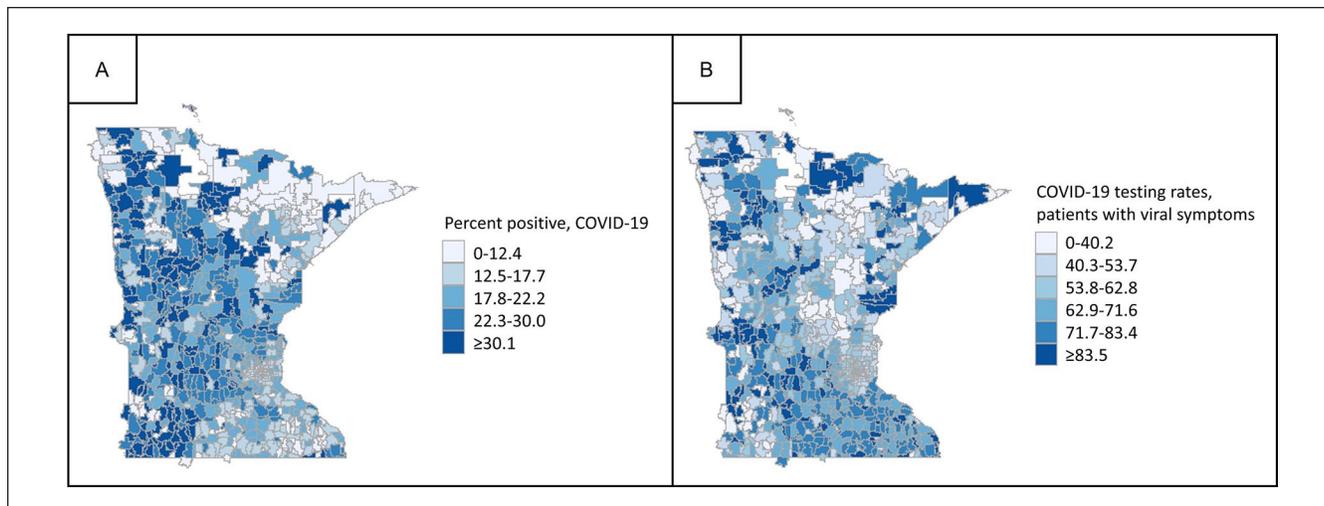
Figure 2. (A) COVID-19 positivity rates among all tests and (B) COVID-19 testing rates among patients with viral symptoms, by zip code tabulation area, Minnesota, *MMWR* week 47, 2020. An *MMWR* (*Morbidity and Mortality Weekly Report*) week is the week of the epidemiologic year used by the Centers for Disease Control and Prevention for disease reporting.

positivity. The program was able to generate and iteratively update findings within weeks of initiation, further illustrating the promise of collaboration across health systems and public health agencies to inform decisions on critical public health issues, such as targeting testing resources.

Although the Consortium's COVID-19 Project is a clear example of the versatility and customizability of EHR data for public health purposes, such collaborations are rare. The National Academy of Medicine recognized the potential for EHR data to improve care for patients across most health conditions and has strongly advocated for the use of EHR data for research, quality improvement, program evaluation, and public health.[21] The National Academy of Medicine emphasized the importance of moving beyond claims data, which can have considerable lag time and often lack data on laboratory results, certain comorbidities, social determinants of health, and demographic characteristics that are commonly available in an EHR.[21,22] Our ability to report on disparities by race and ethnicity, language, and area-level factors after linking with US Census data highlights the potential of EHR data. Our multisystem results were similar to those at Vanderbilt University Medical Center, where disparities in positivity rates by language were noted in a single health system's EHR data, and our model expands on their work by merging data from numerous health systems across the state.[10] Furthermore, using our internal data, we were able to reproduce trends across Minnesota from data collected at state testing facilities and reported to MDH, suggesting our approach had good face validity.

The early successes of the Consortium's COVID-19 project were the result of several factors. First, before the beginning of the pandemic, an initial Consortium working group had recently convened and agreed to collaborate on disparities-focused projects. Second, the use of a federated data approach, with each health system providing summary data and no patient-level identifiable information, was key in overcoming previous reluctance to sharing data across health systems and provided additional security protections for patient data. Third, it was understood that all health systems would, at least initially, be contributing time, resources, and summary data in kind. Remarkably, all partners were committed to the public health benefit that the Consortium could provide and agreed to this stipulation. Fourth, we understood the critical need to include partner organizations. The Institute for Clinical Systems Improvement convenes the chief executive officers of Minnesota's major health systems quarterly as part of its MN Health Collaborative initiatives.[23] MN Community Measurement works with health care systems and providers to collect and analyze clinical data for quality measures.[24] Inclusion of these organizations contributed importantly to the success of the Consortium, in part because of their existing relationships and years of experience facilitating multi–health system collaborations throughout the state.

In addition to COVID-19 syndromic surveillance, the Consortium is working with MDH to enable monitoring of COVID-19 vaccination rates in certain subpopulations. The Consortium will report on vaccination rates by zip code, race and ethnicity, language, age, comorbidities, and area-level factors.[25] This new initiative highlights the flexibility of our methods to quickly adapt to changing public health priorities, in this case, from a focus on COVID-19 testing to vaccination.

Limitations

This study had several limitations. First, the Consortium includes most of Minnesota's large health systems, but

smaller health systems and clinics are not currently participating, as demonstrated by limited coverage in western and rural Minnesota. Smaller health systems without staff resources to participate on their own may be able to do so by leveraging existing data systems used by MN Community Measurement for collecting clinical data for quality measures; thus, efforts are underway to involve more health systems and clinics in the Consortium. Second, the data presented herein include only patients who sought care at the participating health systems; results cannot be extrapolated to people who did not seek care despite experiencing COVID-19–like symptoms. In addition, analyses of summary-level EHR data may overrepresent patients who seek care at multiple health systems or undercount COVID-19 testing in these same patients. Nonetheless, the distribution of racial and ethnic populations in our data are within 1 percentage point of US census estimates for Minnesota. We are currently in the process of implementing secure, deidentified methods for deduplicating patients across systems. Finally, we did not examine severity of illness but have plans to incorporate outcomes such as hospitalization and mortality in future updates.

Practice Implications

In less than 2 months, the Consortium developed a federated data network and reported on patients who received positive COVID-19 test results and who had symptoms of viral illness, by race, ethnicity, and area-level factors across 8 large health systems. This information bridged important gaps in COVID-19 syndromic surveillance for MDH, particularly granular data on patient race, ethnicity, and language, as well as data on patients with symptoms of viral illness who did not receive testing. The Consortium is currently focusing on COVID-19–related surveillance, but the governance, agreements, workflow, data models, and process infrastructure that have been developed to address the COVID-19 pandemic can provide an important foundation for future public health surveillance projects using EHR data. Beyond the current pandemic, the Consortium is well positioned to monitor chronic conditions such as diabetes, hypertension, cardiovascular disease, depression, and opioid use disorder and will be able to react to novel crises that might emerge.

Declaration of Conflicting Interests

Funding

ORCID iD

Tyler N. A. Winkelman, MD, MSc https://orcid.org/0000-0002-9581-5223

Supplemental Material

Supplemental material for this article is available online.

References

1. Gold JA, Rossen LM, Ahmad FB, et al. Race, ethnicity, and age trends in persons who died from COVID-19—United States, May–August 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(42):1517-1521. doi:10.15585/mmwr.mm6942e1

2. Khazanchi R, Evans CT, Marcelin JR. Racism, not race, drives inequity across the COVID-19 continuum. *JAMA Netw Open*. 2020;3(9):e2019933. doi:10.1001/jamanetworkopen.2020.19933

3. Bilinski A, Emanuel EJ. COVID-19 and excess all-cause mortality in the US and 18 comparison countries. *JAMA*. 2020;324(20):2100-2102. doi:10.1001/jama.2020.20717

4. Henry B, Canadian Pandemic Influenza Preparedness (CPIP) Task Group. Canada's pandemic influenza preparedness: surveillance strategy. *Can Commun Dis Rep*. 2018;44(1):14-17. doi:10.14745/ccdr.v44i01a04

5. Institute of Medicine. Emerging technical tools. In: Knobler SL, Mack A, Mahmoud A, Lemon SM, eds. *The Threat of Pandemic Influenza: Are We Ready?* National Academies Press; 2005:254-314.

6. Chan AT, Brownstein JS. Putting the public back in public health—surveying symptoms of COVID-19. *N Engl J Med*. 2020;383:e45. doi:10.1056/NEJMp2016259

7. Bollyky TJ, Patrick SM. *Improving Pandemic Preparedness: Lessons From COVID-19*. Council on Foreign Relations; 2020. Accessed February 7, 2021. https://www.cfr.org/report/pandemic-preparedness-lessons-COVID-19/findings

8. Krieger N, Gonsalves G, Bassett MT, Hanage W, Krumholz HM. The fierce urgency of now: closing glaring gaps in US surveillance data on COVID-19. *Health Affairs* blog. April 14, 2020. Accessed October 6, 2021. https://www.healthaffairs.org/do/10.1377/hblog20200414.238084/full

9. Birkhead GS, Klompas M, Shah NR. Uses of electronic health records for public health surveillance to advance public health. *Annu Rev Public Health*. 2015;36:345-359. doi:10.1146/annurev-publhealth-031914-122747

10. Wilkins CH, Friedman EC, Churchwell AL, et al. A systems approach to addressing COVID-19 health inequities. *NEJM Catalyst*. 2021;2(1). doi:10.1056/CAT.20.0374

11. Aliabadi A, Sheikhtaheri A, Ansari H. Electronic health record–based disease surveillance systems: a systematic literature review on challenges and solutions. *J Am Med Inform Assoc*. 2020;27(12):1977-1986. doi:10.1093/jamia/ocaa186

12. Centers for Medicare & Medicaid Services. Chronic conditions data warehouse: condition categories. Published 2017. Accessed October 6, 2021. https://ccwdata.org/web/guest/condition-categories

13. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision*. Volume 2. World Health Organization; 2004. Accessed September 26, 2021. https://apps.who. int/iris/bitstream/handle/10665/42980/9241546530_eng. pdf?sequence=1&isAllowed=y

14. US Census Bureau. American Community Survey 2015-2019 5-year data release. Published December 10, 2020. Accessed February 7, 2021. https://www.census.gov/newsroom/press-kits/2020/acs-5-year.html

15. US Department of Agriculture Economic Research Service. Rural–urban commuting area codes. Published August 17, 2020. Accessed February 7, 2021. https://www.ers.usda. gov/data-products/rural-urban-commuting-area-codes .aspx

16. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA*. 2020;323(11):1061-1069. doi:10.1001/jama.2020.1585

17. Guan WJ, Ni ZY, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. 2020;382(18):1708-1720. doi:10.1056/NEJMoa2002032

18. Guan WJ, Liang WH, Zhao Y, et al. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. *Eur Respir J*. 2020;55(5):2000547. doi:10.1183/13993003.00547-2020

19. Tian S, Hu N, Lou J, et al. Characteristics of COVID-19 infection in Beijing. *J Infect*. 2020;80(4):401-406. doi:10.1016/j. jinf.2020.02.018

20. Grant MC, Geoghegan L, Arbyn M, et al. The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): a systematic review and meta-analysis of 148 studies from 9 countries. *PLoS One*. 2020;15(6):e0234765. doi:10.1371/journal.pone.0234765

21. Okun S, McGraw D, Stang P, et al. *Making the Case for Continuous Learning From Routinely Collected Data*. Institute of Medicine; 2013. Accessed February 7, 2021. https://nam. edu/perspectives-2013-making-the-case-for-continuous-learning-from-routinely-collected-data

22. Wilson J, Bock A. *The Benefit of Using Both Claims Data and Electronic Medical Record Data in Health Care Analysis*. Optum; 2012.

23. Institute for Clinical Systems Improvement. MN Health Collaborative. Accessed February 7, 2021. https://www.icsi. org/programs/mn-health-collaborative

24. MN Community Measurement. About us. Accessed February 7, 2021. https://mncm.org/about

25. Centers for Disease Control and Prevention, Agency for Toxic Substances and Disease Registry. CDC/ATSDR Social Vulnerability Index. Updated April 28, 2021. Accessed September 17, 2021. https://www.atsdr.cdc.gov/ placeandhealth/svi/index.html