
Research and Applications

Building capacity of community health centers to overcome data challenges with the development of an agile COVID-19 public health registry: a multistate quality improvement effort

Lisa Romero¹, Pedro B. Carneiro², Catharine Riley¹, Hollie Clark¹, Raymonde Uy², Michael Park¹, Tebitha Mawokomatanda¹, Jennifer M. Bombard¹, Alison Hinckley¹, and Julia Skapik²

¹Centers for Disease Control and Prevention COVID-19 Response, Health Systems and Worker Safety Task Force, Atlanta, Georgia, USA, and ²National Association of Community Health Centers, Department of Clinical Affairs, Bethesda, Maryland, USA

Corresponding Author: Lisa Romero, DrPH, MPH, Division of Reproductive Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, 4770 Buford Highway, MS S107-2, Atlanta, GA 30341, USA; eon1@cdc.gov

Received 18 May 2021; Revised 2 September 2021; Accepted 11 October 2021

ABSTRACT

Objective: During the coronavirus disease 2019 (COVID-19) pandemic, federally qualified health centers rapidly mobilized to provide SARS-CoV-2 testing, COVID-19 care, and vaccination to populations at increased risk for COVID-19 morbidity and mortality. We describe the development of a reusable public health data analytics system for reuse of clinical data to evaluate the health burden, disparities, and impact of COVID-19 on populations served by health centers.

Materials and Methods: The Multistate Data Strategy engaged project partners to assess public health readiness and COVID-19 data challenges. An infrastructure for data capture and sharing procedures between health centers and public health agencies was developed to support existing capabilities and data capacities to respond to the pandemic.

Results: Between August 2020 and March 2021, project partners evaluated their data capture and sharing capabilities and reported challenges and preliminary data. Major interoperability challenges included poorly aligned federal, state, and local reporting requirements, lack of unique patient identifiers, lack of access to pharmacy, claims and laboratory data, missing data, and proprietary data standards and extraction methods.

Discussion: Efforts to access and align project partners' existing health systems data infrastructure in the context of the pandemic highlighted complex interoperability challenges. These challenges remain significant barriers to real-time data analytics and efforts to improve health outcomes and mitigate inequities through data-driven responses.

Conclusion: The reusable public health data analytics system created in the Multistate Data Strategy can be adapted and scaled for other health center networks to facilitate data aggregation and dashboards for public health, organizational planning, and quality improvement and can inform local, state, and national COVID-19 response efforts.

Key words: SARS-CoV-2, COVID-19, health centers, EHR data, public health informatics infrastructure

INTRODUCTION

The coronavirus disease 2019 (COVID-19) pandemic poses serious health risks to the U.S. population, but particularly for communities and populations with long-standing social inequities and health disparities that increase the risk for COVID-19 infection, severe illness, and death.^{1,2} The Health Resources and Services Administration (HRSA) Health Center Program supports nearly 1400 federally qualified health centers that provide comprehensive primary health care to approximately 30 million patients at more than 13 000 service sites across the United States.³ Health centers are often located in areas where economic, geographic, or cultural barriers limit access to affordable health care and serve as a critical safety net for populations at high risk for COVID-19, including racial and ethnic minority groups, migrant and agricultural workers, people experiencing homelessness, residents of public housing, and veterans.^{3,4} Furthermore, health centers also serve essential nonhealth-care workers in industry sectors (ie, food processing, manufacturing, construction, health aides) that are at high risk for COVID-19 and disproportionately impacted by the pandemic.⁵⁻⁷ Throughout the COVID-19 pandemic, health centers have supported SARS-CoV-2 testing, COVID-19 follow-up care, and vaccination for these populations served. During this time, many health centers have reached communities beyond their usual patient populations, to reduce the health inequities exacerbated by the pandemic; however, the extent of such outreach has not been measured.⁸⁻¹¹

To address the constraints introduced by the pandemic, health centers pivoted many of their services and programs to focus on crisis operations and continuity of patient care. There was an urgent need for more timely and comprehensive public health data to guide decision-making. However, there were also data collection and reporting challenges affecting health centers. This stemmed from limits in the health information technology (HIT) infrastructure at the health center level and challenges with public health data extraction (ie, development of comprehensive reports, lack of medical terminology) and lack of alignment at the local and state levels.¹² In addition, early in the pandemic COVID-19 was a novel medical condition without formal concepts in existing medical terminologies and adequate guidance to support data capture and billing. Although the informatics community rallied to provide novel concepts and codes, it took time for teams to implement them into electronic health records (EHRs) and for providers to become familiar with each code and its appropriate use (eg, the availability of terminology codes did not immediately align with the availability of these codes in HIT systems). Throughout the first year, the COVID-19 data response was hampered by historic and prevalent challenges with data quality, interoperability, and standardization.

Many health centers partner with regional Health Center-Controlled Networks (HCCNs), HRSA-funded networks that help health centers improve quality of care and patient safety by using HIT to reduce costs and improve care coordination, and state Primary Care Associations (PCAs), HRSA-funded nonprofit organizations that provide training and technical assistance to help health centers improve programmatic, clinical, and financial performance and operations. Yet, a national infrastructure above the HCCNs and PCAs to support the sharing of patient-level health center data for public health evaluation and ad-hoc analytics across the United States did not exist prior to the COVID-19 pandemic. The public health emergency revealed that a national health center data infrastructure that allows for standardization, sustainable data extraction, and analysis for continuous data quality improvement was

needed to describe and monitor the health burden, disparities, and overall COVID-19 impact on populations served by health centers.

To respond to this need, and as part of the Centers for Disease Control and Prevention (CDC) COVID-19 emergency response, CDC, in partnership with the National Association of Community Health Centers (NACHC), and in coordination with HRSA, aligned strategies and resources to support health centers' efforts to track the response and impact of COVID-19. NACHC, the national association dedicated to the support of community-based health centers and the expansion of health care access for the medically underserved and uninsured, facilitates efforts to build health centers' capacity to effectively respond to COVID-19 and future public health emergencies through a centralized technical infrastructure that provides for standardized data flow, content, and technical assistance. Through this partnership, a Multistate Data Strategy was implemented to inform continuous improvements for COVID-19 care with innovative, collaborative, and measurable models to enhance health centers' response efforts.

The Multistate Data Strategy developed a foundation for a reusable infrastructure and data sharing procedures for health centers and public health partners. The primary objectives of the approach included the following: (1) support activities in multiple state and regional HCCNs and PCAs across the United States to implement an informatics-enabled, public health reporting project to address the impact of COVID-19 through clinical data reuse; (2) build an infrastructure for data aggregation, data quality improvement, and innovation in HIT; (3) collaborate with partners to support the utilization of existing data and systems for data collection to support state and local health entities with monitoring the health burden, disparities, and overall impact of COVID-19 on populations served by health centers; and (4) gather and analyze data in a health center and NACHC-managed registry on COVID-19 for quality improvement, public health surveillance, and outcomes evaluation to further inform local, state, and national COVID-19 response efforts.

This paper describes the development of a reusable public health data analytics system for data aggregation and quality improvement developed by NACHC; reports on the project's feasibility, readiness, data challenges, and quality improvement efforts; and discusses key lessons learned that can be adapted and scaled up for other health center networks in coordination with public health, hospitals, and community-based organizations to inform local, state, and national COVID-19 response efforts and more effectively respond to future public health emergencies.

MATERIALS AND METHODS

Multistate Data Strategy partner selection

Through an open and competitive request for proposals, 6 regional and state HCCNs and PCAs were selected, with a goal of reaching diverse geographic regions and populations, including urban and rural populations, low-income and uninsured individuals, racial and ethnic minority groups, sexual and gender minority groups, refugees, and individuals with housing insecurity and homelessness. The Multistate Data Strategy project partner locations represent states with HCCN or PCA health center members (Figure 1).

Assessment of health center data challenges

The NACHC Informatics Team utilized a human-centered design framework¹³ (Figure 2) to engage the project partners and direct the scope of the work, assess data challenges and goals, use patient

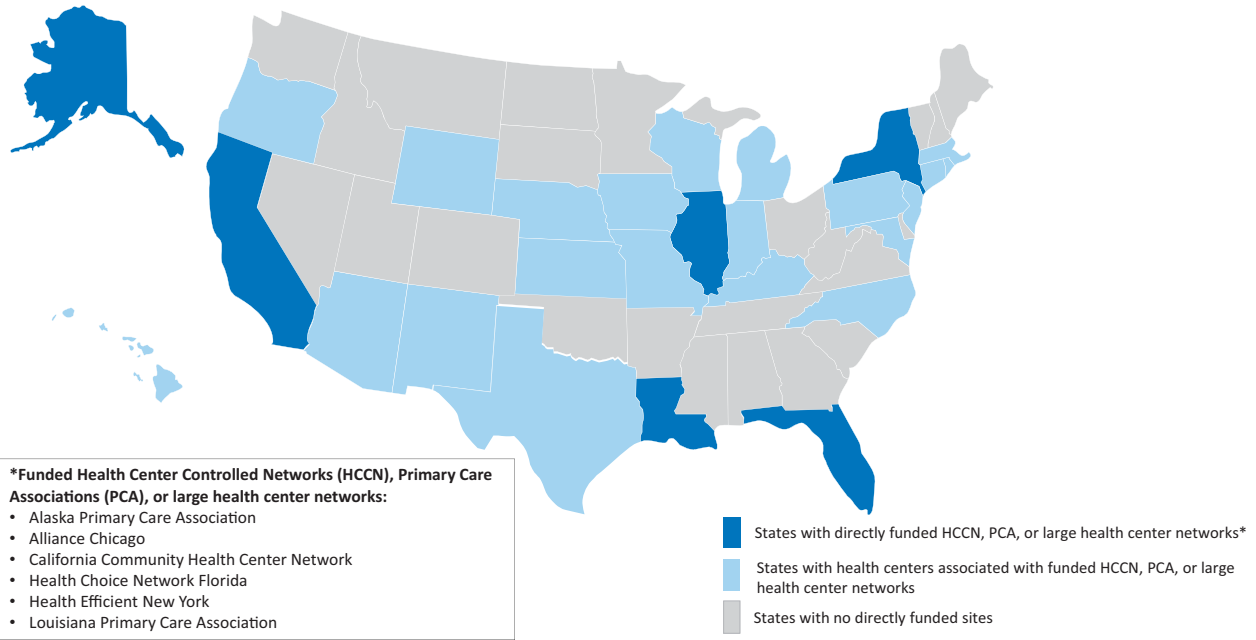


Figure 1. Multistate Data Strategy Project Partners. This figure indicates where the Multistate Data Strategy project partners Health Center-Controlled Networks (HCCN) and state or regional Primary Care Associations (PCAs) are located.

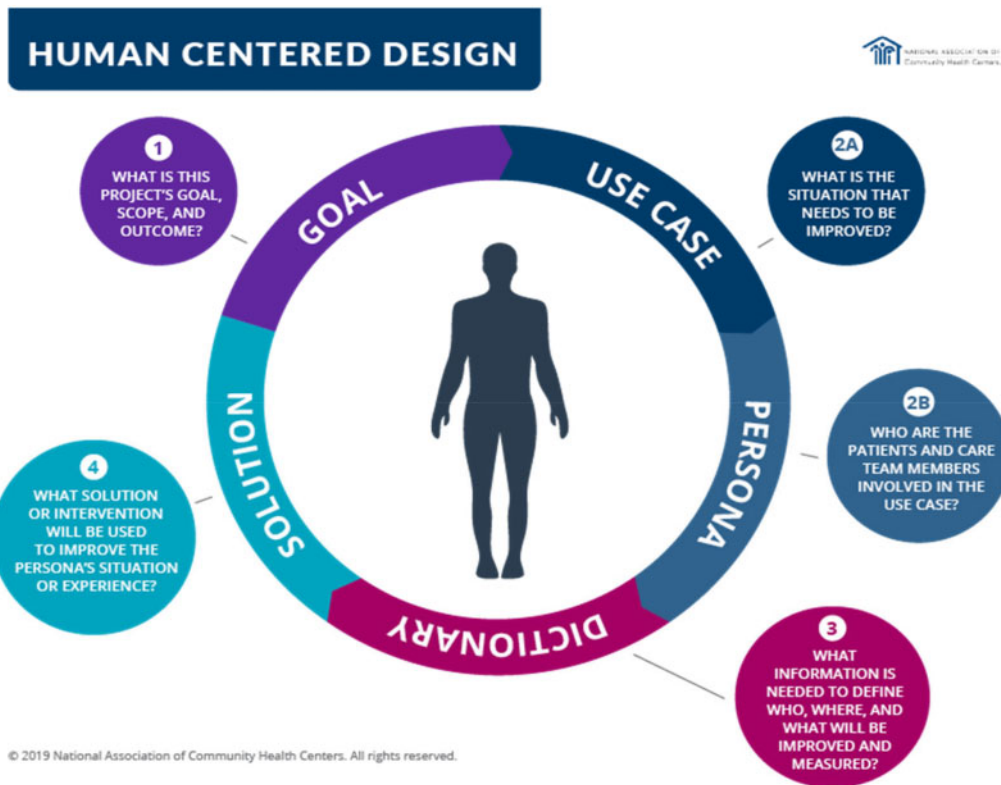


Figure 2. Human-centered design framework applied to the Multistate Data Strategy. This figure indicates the formal steps of the human-centered design framework used to build materials and conduct our Multistate Data Strategy product activities.

stories, and identify shared goals and relevant use cases for alignment. Project partners were asked to document data challenges, workflows, modifications to business practices, and narratives contextualizing these challenges into patient stories. The NACHC Infor-

matics Team collected the qualitative information and summarized each project partners assessment data. This information served as the foundation for the COVID-19 registry design and proposed measures. A formal use case evaluation was conducted. The use

cases considered multiple perspectives and sets of requirements, including dashboards for health center response and planning, clinical care, and prevention activities for COVID-19, contact tracing, and public health reporting.

Development of informatics-informed public health data architecture

Cloud-based data infrastructure

To support the Multistate Data Strategy, the NACHC Informatics Team established a cloud-based data infrastructure instance (ie, virtual server) to allow for the intake and processing of a variety of data formats and sources asynchronously, with variable computing capacity and with redundancy to protect data loss. Data in the initial phase were submitted by project partners and the NACHC Informatics Team manually uploaded and imported to the systems data lake (ie, storage repository that holds a vast amount of raw data). Future iterations of the system will allow for direct application programming interface (API) data transfers. Uploaded data were cataloged and metadata were captured in a relational database used to clean, validate, merge, and collate information through automated processes primarily written in Java and SQL. The code is available for view at <https://github.com/NACHC-CAD>. The streamlined data were then published into a data warehouse. The NACHC Informatics Team applied the Automated Extract, Transform and Load (ETL) processes to the raw data to normalize and move it into a data warehouse, which in turn fed downstream information systems (Figure 3).

COVID-19 data dictionary and data elements

The NACHC Informatics Team developed a data dictionary containing data elements and concepts expected to be available in EHRs and other electronic systems (eg, data warehouses, population health sys-

tems, inventory management systems). Data elements were categorized into 12 COVID-19 domains: demographics (eg, race/ethnicity, education, occupation), social determinants of health (SDOH; eg, housing, insurance, income, material security, transportation), administrative/clinic level data (eg, facility, zip code), conditions/diagnosis (eg, cardiovascular disease, hypertension, diabetes, obesity, suspected COVID-19 infection), encounters (eg, hospitalization, inpatient diagnosis, telemedicine), equipment (eg, face mask), exposures (eg, known COVID-19 exposure, essential worker, homeless), interventions (eg, isolation, case investigation, contract tracing), observations (eg, at-risk for COVID-19, need for evaluation from medical professional), signs and symptoms (eg, fever, cough, difficulty breathing), testing (eg, type of diagnostic test, antibody test), and vital signs (eg, height, weight, fever). Subsequently, categories for therapeutics and vaccination were incorporated as those elements became available to health centers. Domains and element definitions were based on the best content available at the time of the COVID-19 pandemic, with the goal of harmonizing data requirements with existing health center data, such as the HRSA Uniform Data System (UDS),¹⁴ state-level data exchange definitions for immunizations published by CDC,¹⁵ the Office of the National Coordinator for Health Information Technology (ONC) standards-based content (eg, U.S. Core Data for Interoperability [USCDI]),¹⁶ NACHC's Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences (PRAPARE)¹⁷ and the availability of existing value sets in the Value Set Authority Center of the National Library of Medicine.¹⁸ For COVID-19-specific content, we referenced the data elements and value sets created through the COVID-19 Interoperability Alliance,¹⁹ which made content rapidly available to systems and users during the pandemic. Informatics staff at the project partner sites responsible for data extraction conducted an organizational data environmental scan to map data elements within each domain and provided feedback and suggested additional data elements.

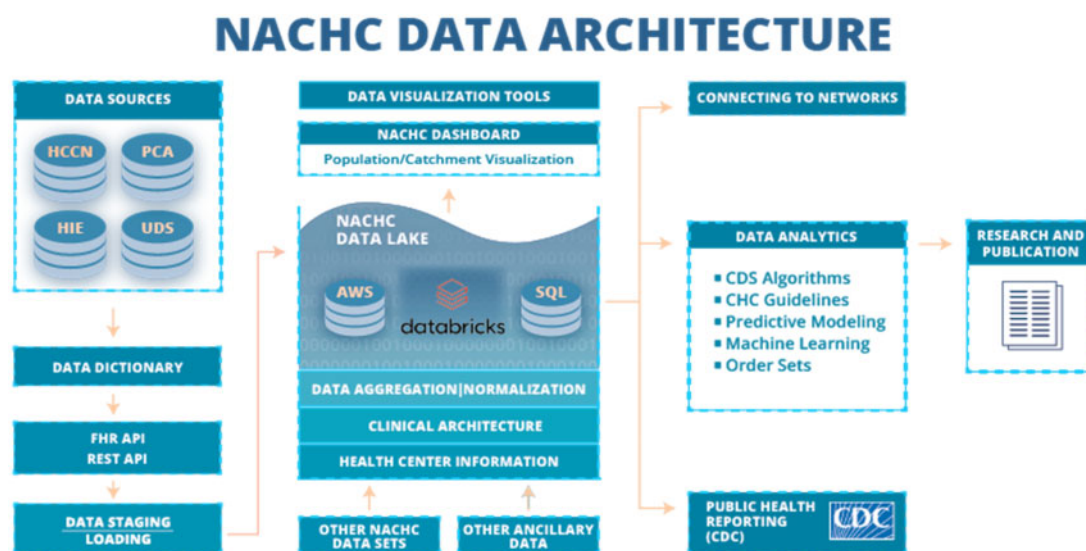


Figure 3. Multistate Data Strategy COVID-19 Data Architecture. Diagram illustrates the conceptual approach to NACHC's data architecture framework to consolidate and standardize data from multiple sources in multiple formats. Process describes ingestion of data from various sources, the building of a data dictionary, and the process for deploying data into a data lake, where data are standardized and transformed. The process ends with the data being used for various solutions, once it is properly defined. AWS: Amazon API Gateway; API: application programming interface; CDC: Centers for Disease Control and Prevention; CDS: Clinic Decision Support; NACHC: National Association of Community Health Centers; CHC: Community Health Centers; FHR: Fast Healthcare Interoperability Resource; HCCN: Health Center-Controlled Network; HIE: Health Information Exchange; PCA: Primary Care Association; REST: Representational State Transfer; SQL: Structured Query Language; UDS: Uniform Data System.

Quality improvement efforts

The project partners were asked to improve the data dictionary by having their data teams review the content, provide feedback, perform a formal feasibility assessment using an internally developed instrument, and perform queries at an aggregate level across their EHRs to validate whether assumptions made about the data were correct. Project partners also were asked to perform formal chart reviews of patients for quality assurance of data and to identify opportunities to close data gaps and improve mapping of data to the data dictionary.

Data extraction and transfer

The Multistate Data Strategy leveraged existing data collection approaches from health centers. Given all project partners were primary or secondary data users and aggregators of clinical data warehouses, NACHC engaged the respective HCCN and PCA project partners in separate data use agreements to extract and provide deidentified data for public health purposes. Project partners transmitted data only after required data-sharing agreements were completed among NACHC and project partners. NACHC received this transfer of deidentified data to the COVID-19 NACHC registry. NACHC asked partners to provide dummy patient identifiers for their patients so multiple data extracts can be aggregated while representing each individual patient's data longitudinally for evaluation of patient outcomes. All data sent to NACHC are stored in a secured cloud data warehouse. Secure user access to the cloud data is limited and password protected.

Analysis of extracted data and information for quality improvement

Descriptive statistics were used to report the aggregate-level descriptive content (eg, counts, percentages) of the initial extracted data. Information from the project partners was reviewed by the NACHC Informatics Team to describe the processes, identify data challenges, and document quality improvement efforts. Dashboards and sample data were shared with partners for validation and to provide preliminary analytics as the project moved into the monthly data phase and updated routinely. This activity was reviewed by CDC and was conducted consistent with applicable federal law and CDC policy (see eg, 45 C.F.R. part 46, 21 C.F.R. part 56; 42 U.S.C. §241(d); 5 U.S.C. §552a; 44 U.S.C. §3501 et seq.).

RESULTS

Findings of assessment of health center data challenges

Between August 2020 and March 2021, project partners completed the initial data feasibility and aggregation phase and transitioned to extract and report deidentified patient-level data. The patient-level data were used to examine patient outcomes, such as aggregate number of COVID-related diagnosis and volume of testing, and to evaluate the impact of COVID-19 on the health centers and the populations they served. The initial assessment highlighted that many health centers could not rapidly pivot to address COVID-19 public health data capture and sharing because of the constraints by or on their EHR vendor products' ability to extract data, limited technical staff, and data challenges across partner health centers, including the technical and operational limitations of data sharing arrangements. Many challenges identified by health centers were not related to the COVID-19 pandemic but were more often related to common HIT interoperability challenges, including misaligned local, state,

and federal reporting requirements and formats (eg, varying terminologies, subjective definitions, aggregate vs patient-level data), lack of a unique patient identifier, lack of comprehensive access to pharmacy, claims and laboratory data, large volumes of missing data, proprietary data standards such as lab and medication vendor terms, and externally controlled extraction methods such as third-party data services or EHR charges for extraction.

To address the identified challenges, project partners initiated quality improvement efforts to optimize the data capture and data-flow. Quality improvement efforts were 3-fold: (1) the clinical and public health activities around the pandemic; (2) the data validation and cleaning activities to improve data quality; and (3) the organization-designed innovation project to address interoperability challenges. All project partners provided feedback on the project data regularly to health center partners and worked to make improvements to the EHR and other HIT systems to support more rapid and flexible data entry and analytics, including actions to address gaps in data capture, integrate data streams, implement electronic case reporting, gain access to external data sources, and use HIT tools (eg, telehealth, scheduling, population health) to support health center management of patients affected by COVID-19.

Partner HIT profile and initial sample characteristics

In total, the 6 project partners—4 HCCNs and 2 state PCAs reported over 900 000 patients served through approximately 3.6 million medical encounters between January and November 2020 (Table 1). Nearly 350 000 SARS-CoV-2 tests were reported, which accounted for an average 10% positivity rate across partners. About 13% of the initial population had a diagnosis code (ie, ICD-10 Code) associated with COVID-19 exposure or infection. Project partners varied in size, population distribution, and level of access to EHR systems. In total, only 4 of the 6 project partners reported having direct access to their EHR data, while the 2 state PCAs managed the EHR data using a third-party data vendor because of variation in the use of different EHRs across the state. Further, access to external data sources varied across states and type of organization. Two project partners reported being able to import external EHR

Table 1. Multistate Data Strategy project partner initial data extraction from EHRs, January to November, 2020

Partner HIT profile (<i>n</i> = 6)	
EHR data	
Direct access	4 (66%)
Managed at health center level	2 (33%)
EHR health record exchange capability	
Import external records within the same vendor	2 (33%)
Import external records from any vendor	1 (16%)
Unable to electronically import external records	3 (50%)
Initial data extraction profile	
Patient population	794 572
% Patients with COVID diagnosis ^a	13%
Medical encounters	2 846 566
SARS-CoV-2 tests	346 503
% Test positive ^b	12%

^aUnique patients with an associated COVID-related diagnosis code.

^bTotal numbers of SARS-CoV-2 tests performed. These data were extracted and aggregated by the Multistate Data Strategy project partners from EHR data of participating health centers.

HIT: health information technology; EHR: electronic health record; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2.

data (eg, transfer medical records across providers) if the data sources were from the same EHR vendor, one partner reported they could import external EHR data from any vendor, and the remaining reported being unable to do so. Although weekly data pulls would have been preferred by the NACHC Informatics Team, the project had to utilize monthly pulls because of the effort involved and limited bandwidth during the pandemic.

EHR environmental scan

Among the 6 project partners, all reported lack of access to all or some of the data elements across all domains (Table 2). Across all domains, project partners reported that significant portions of the data were not available in a structured format (eg, text existed only within a medical note or in paper form as a scanned document). Although data capture may have been possible with additional effort, project partners reported that the data elements were not collected in full for all patients and a high number of missing data elements were reported for some domains. In general, organizations had high level of access to demographic and clinic-provided medical data but less access to other data types and a high proportion of missing data around SDOH and utilization data. While many partners had some access to electronic health information exchange (HIE) data, it was not necessarily easily integrated into the local record even if it were accessible to clinic staff.

All project partners reported challenges related to data capture and use across all domains. No project partners reported being able to capture data in a structured format or extract data related to equipment (eg, personal protective equipment) or inventory (eg, vaccine). However, some project partners reported they were in the process of exploring or obtaining supply chain management software to address this challenge. Additionally, project partners reported that several key data elements, such as SDOH (eg, employment, economic stability) and exposure (eg, industry/occupation), were inconsistently collected due in part to lack of standardization of the data or incorporation into medical workflow, resulting in missing data.

Several project partners reported that EHR systems lacked the ability to extract data using federally mandated terminology mappings (eg, Logical Observation Identifiers Names and Codes [LOINC] and Prescription for Electronic Drug Information Exchange [RxNorm]).^{20,21} Further, all project partners reported some use of standard medical terminology standards and access to International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM),²² Current Procedural Terminology (CPT),²³ and LOINC²⁴ codes; however, only one project partner had system access to RxNorm and Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT), and specific codes frequently were not mapped resulting in data that were not easily extractable. Lastly, interoperability between health centers' EHR systems and state-level systems (eg, immunization registries and Health-Information Exchanges) varied, but was generally difficult. For example, project partners identified the data challenge of reporting SARS-CoV-2 test type, including inconsistent use of numeric values or code applied to the test names and missing standard code descriptions (ie, LOINC; Figure 4). Furthermore, project partners submitted the monthly data extracts in different data formats (ie, character, factor, numeric), which further created challenges for data collection, cleaning, and aggregation.

DISCUSSION

The Multistate Data Strategy was developed to support the complex interoperability needs of the project partners' health systems data infrastructure in the context of the need to respond rapidly to the COVID-19 pandemic.²⁵ During the initial assessment of implementing an informatics-enabled public health reporting system, our findings revealed that despite the presence of EHR data standards,²⁵ project partners were limited in their ability to easily extract all of the relevant data categories mapped into medical terminologies. This gap in mapping created a significant burden and required manual data processes to achieve data conformant to the COVID-19

Table 2. Environmental scan of EHR data from Multistate Data Strategy project partners organizations

	Mean proportion of elements within domains available at EHR	Percentage of available data elements that exist within a structured data field
Domains (<i>n</i>)		
Demographics (14)	77%	98%
SDOH (21)	94%	98%
Administrative (7)	75%	43%
Conditions/diagnosis (27)	70%	69%
Encounter (11)	35%	29%
Equipment (5)	0%	—
Exposure (10)	12%	75%
Intervention (10)	20%	40%
Observation (2)	38%	75%
Signs and symptoms (33)	41%	72%
Testing (44)	53%	77%
Vital signs (4)	88%	63%
Terminology supported (<i>n</i> = 6)		
ICD-10-CM	6 (100%)	
CPT	6 (100%)	
LOINC	6 (100%)	
RxNorm	1 (16%)	
SNOMED-CT	1 (16%)	

EHR: electronic health record; SDOH: social determinant of health; ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM); CPT: current procedural terminology; LOINC: Logical Observation Identifiers Names and Codes; RxNorm: Prescription for Electronic Drug Information Exchange; SNOMED-CT: Systematized Nomenclature of Medicine Clinical Terms.

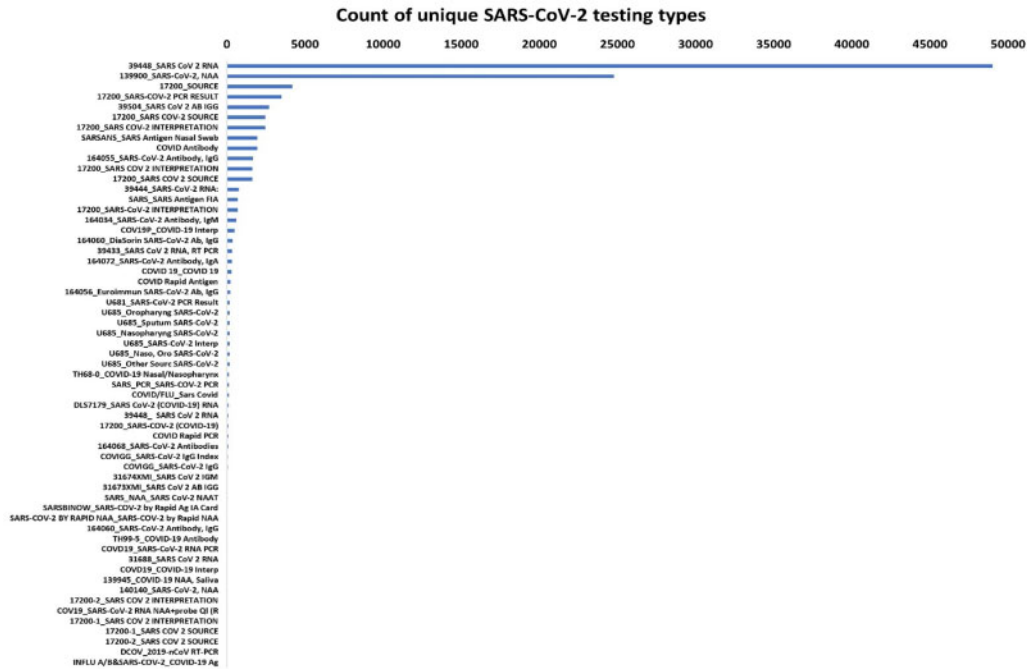


Figure 4. Example of numeric values or code applied to the SARS-CoV-2 test names. This example highlights the Multistate Data Strategy project partners data challenge of reporting SARS-CoV-2 test type, including inconsistent use of numeric values or code applied to the test names and no use of standard code descriptions. SARS-CoV-2: severe acute respiratory syndrome coronavirus 2.

data dictionary. Further, although federal interoperability data standards have been in place for nearly a decade and additional requirements are in process, data from health centers often lacks standardization and completeness. This remains a significant barrier to real-time data analytics and efforts to improve health outcomes and mitigate inequities through data-driven responses.

Standardization and mapping of EHR content

Implementation of the Multistate Data Strategy highlights the importance of assuring data quality and comprehensive content of clinical care, data capture and extraction for public health data sharing, and may have broader implications to patient care data exchange requirements across systems to integrate regional healthcare system data to support public health efforts. Despite the effort by the COVID-19 Interoperability Alliance to develop and publish novel SARS-CoV-2 and COVID-19 content rapidly to the health care industry, project partners lacked access to standardized data needed to achieve high-quality COVID-19 data extracts. The National Library of Medicine’s Value Set Authority Center rapidly published COVID-19 value sets¹⁸ to provide opportunities to standardize EHR coding. However, while project partners EHRs’ had access to ICD-10-CM codes to identify patients with specific diagnoses and clinical problems, most project partners lacked the mappings to SNOMED-CT to identify coded, structured clinical findings and tests linked to their appropriate LOINC code, which limited data quality and required the NACHC Informatics Team to recategorize text-based data without the ability to perform additional validation. Our project partners reported that EHR vendors provided limited or no direct support for mapping of medical terminology despite that there are few to no terminologists employed in health center settings; this lack of support from the system itself for standardized data extraction had the greatest impact on this project and indicates

that despite federal certification standards, many customers do not functionally have some of the most basic health IT standards. Greater oversight and increased surveillance and test complexity from federal regulatory bodies could have a significant impact on vendor responsiveness to the urgent need for standardized data extraction. Lack of standardization was also limited by systems that allowed hand entry of conditions, tests, medications, and test results, by misaligned state and local supplies and lack of guidance and training of care teams on proper code use and coding.

While third-party systems or population health tools may make analytics easier for users at the health center level, large-scale mapping of content to an often-proprietary set of new definitions slowed data transmission and was often at a significant cost. Although some project partners were able to support mapping efforts locally or use contractors, others relied on NACHC Informatics Team-sponsored terminology services to organize and map their EHR content. To support the rapid availability of clinical data for both clinical quality improvement and public health, it is critical to deploy, automate, and streamline mapping services to serve EHRs administrators and users, which in turn may improve data quality and ultimately patient care. Furthermore, ensuring that EHR vendors create and sustain mapping tools and that products are equipped with appropriate terminology support and that health centers are staffed with terminology-trained staff may also improve adherence to data standards.

Implementation of the Multistate Data Strategy highlights limitations of EHR systems related to the ability to identify emerging signs of a new communicable condition using symptom-based criteria (eg, suspected COVID infection) and nonmedical risk factor categories (eg, recent travel, occupation, multigenerational home). These data elements are inconsistently captured or captured as unstructured elements and therefore difficult to analyze at the local, state, and national levels. To address this challenge, health centers may consider changes to data capture workflows; the development

of automated extraction and interfaces to improve data reuse by public health stakeholders; greater participation of patients to collect and validate their own structured data as part of the encounter (eg, via tablet or phone applications); the development of or negotiated access to data extraction techniques; and the ability to create structured elements or parse structured data from unstructured data in the EHR. These strategies can maximize the ability to understand the natural course of disease, capture relevant metadata, evaluate patient outcomes, and build a public health digital infrastructure. For example, the NACHC Informatics Team identified nonmedical essential worker status as a significant predictor of risk but were unable to extract these data from EHR systems. Federal regulators and state-level data exchange entities should consider the inclusion of standardized values defining labor-related categories (eg, Department of Labor) in EHR products and provide guidance on use to develop effective essential worker and industry and occupational risk categories to further support coding for COVID-19 and occupational-associated conditions.²⁶

Finally, we identified a need for improved collection of demographics and SDOH data to describe and monitor health disparities in populations and communities at risk for COVID-19. Long-standing social inequities and health disparities have resulted in increased risk for COVID-19 infection, severe illness, and death among racial and ethnic minority populations.^{1,27-29} Our project partners reported challenges capturing and operationalizing SDOH data. Data completeness may be improved if data from patient records and external sources could be collated for data exchange and sharing, workflows were redesigned around SDOH components, or if comprehensive data capture and coding aligned with billing and reimbursement. Standardization and testing of SDOH data across EHRs and population health systems and integration of SDOH content into federal standards and value sets to support organizations and their corresponding technology systems may improve SDOH data capture.^{17,30}

The implementation of this Multistate Data Strategy included 6 project partners and may not be generalizable to all health centers across the United States; however, there were several key lessons learned. Had a public health infrastructure for health centers and their public health partners been in place at the start of the pandemic, critical data from populations and communities at high risk for COVID-19 could have been normalized and quickly analyzed (eg, in days or weeks instead of months) to augment the response. Our efforts to rapidly build and deploy a structured, cloud-based infrastructure to support the public health response to the COVID-19 pandemic for culturally diverse and low-resourced health centers demonstrated it is feasible to rapidly stand up a responsive community of safety-net primary care organizations. Further, many of the challenges faced by project partners in aggregating and extracting data for the Multistate Data Strategy were not specific to the COVID-19 pandemic, but instead represented ongoing challenges with the lack of consistent and comprehensive data capture and extraction from EHRs. HCCNs and PCAs provide a valuable service to their health center partners given that health centers generally do not have the resources to staff the informatics and health IT teams needed to overcome these challenges at the local level. Project partners were often constrained by competing and misaligned local, state, and federal requirements, inadequate funding, and EHRs that lacked the ability to standardize data in federally mandated medical terminologies. Nevertheless, the deployment of new interoperability standards (eg, Fast Healthcare Interoperability Resources [FHIR]) and efforts of the federal government to improve data harmonization among health delivery organizations and technology vendors

may soon yield additional progress toward the development of a national health center data infrastructure for standardization, sustainable data extraction, and analysis for continuous data quality improvement. Although we describe some of the health center data challenges, these challenges also occur in other health systems and represent an opportunity to address an industry-wide set of challenges that limit data usability and reuse.

CONCLUSION

During the COVID-19 pandemic, health centers rapidly mobilized to provide SARS-CoV-2 testing, COVID-19 care, and vaccinations to populations at increased risk for COVID-19 infection, illness, and death. Through the Multistate Data Strategy, a reusable public health data analytics system to aggregate and analyze health center data were developed to address challenges with data capture and sharing procedures between health centers and public health agencies and move toward real-time public health data analytics to describe and monitor the health burden, disparities, and impact of COVID-19 on populations served by health centers. Lessons learned highlight the need for health centers to have HIT systems that support on demand data extraction, to implement best practices to ensure high-quality data capture, and to use data to drive quality improvement in patient care. These changes would enable quality improvement activities and care coordination that can accelerate the work to close care gaps and more effectively respond to public health needs like preventive care and emergencies as well as chronic conditions. Federal partners can play an enabling role in working with clinical and public health communities to provide leadership to help guide the local, state, and regional work and to assess and guide the development of public health informatics infrastructure.³¹ Public health partners can align local, state, federal, and industry data requirements, support high-quality data capture, and provide timely feedback to health centers to guide decision-making. Greater integration of HIT and clinical care through the routine use of standardized medical terminologies for data capture and extraction, implementation of FHIR technology, and stricter compliance to federal guidelines by vendors and consumers of EHR software are key steps that can be taken to best prepare our healthcare systems for future emergency responses. Further, the infrastructure can be adapted and scaled up in other clinical domains across additional health center networks in coordination with public health, hospitals, and community-based organizations to inform local, state, and national COVID-19 response efforts and to help health centers build their capacity to more effectively respond to future public health emergencies.

FUNDING

This project was funded by the US Centers for Disease Control and Prevention.

AUTHOR CONTRIBUTIONS

LR, PBC, CR, HC, RU, MP, and JS contributed to study design, data collection, analysis and interpretation, and manuscript preparation. TM, JMB, and AH contributed to analysis and interpretation and manuscript preparation.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY

The data underlying this article cannot be shared publicly due to ownership by individual health centers serving medically underserved or uninsured, minority, and marginalized communities across the United States. These health centers have different policies, and agreements around data sharing to protect these communities, and have agreed to participate in this project with the condition that their data be shared only in aggregate in order to further protect the communities they serve.

DISCLAIMER

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

ACKNOWLEDGMENTS

We would like to acknowledge the following individuals and organizations who contributed to various aspects of this project: Kathy McNamara, Ron Yee, Russell Brown, Sonia Gil, and Caryn Bernstein (National Association of Community Health Centers); Alaska Primary Care Association; AllianceChicago; California Community Health Center Network; Health Choice Network Florida; Health Efficient New York; Louisiana Primary Care Association.

REFERENCES

- Centers for Disease Control and Prevention. Health Equity Considerations and Racial and Ethnic Minority Groups. 2020. <https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html> Accessed April 2, 2021.
- Wortham JM, Lee JT, Althomsons S, *et al*. Characteristics of persons who died with COVID-19—United States, February 12–May 18, 2020. *MMWR Morb Mortal Wkly Rep* 2020; 69 (28): 923–9.
- Health Resources and Services Administration. HRSA Health Center Program. 2020. <https://bphc.hrsa.gov/sites/default/files/bphc/about/health-centerfactsheet.pdf> Accessed April 2, 2021.
- Health Resources and Services Administration. 2019 Uniform Data System—National Health Center Data. 2020. <https://data.hrsa.gov/tools/data-reporting/program-data/national> Accessed April 2, 2021.
- Dooling K, Marin M, Wallace M, *et al*. The advisory committee on immunization practices' updated interim recommendation for allocation of COVID-19 vaccine—United States, December 2020. *MMWR Morb Mortal Wkly Rep* 2021; 69 (5152): 1657–60.
- Centers for Disease Control and Prevention. Interim List of Categories of Essential Workers Mapped to Standardized Industry Codes and Titles. 2021. <https://www.cdc.gov/vaccines/covid-19/categories-essential-workers.html> Accessed March 22, 2021.
- Selden TM, Berdahl TA. Risk of severe COVID-19 among workers and their household members. *JAMA Intern Med* 2021; 181 (1): 120–2.
- Romero L, Pao LZ, Clark H, *et al*. Health center testing for SARS-CoV-2 during the COVID-19 pandemic—United States, June 5–October 2, 2020. *MMWR Morb Mortal Wkly Rep* 2020; 69 (50): 1895–901.
- Demeke HB, Pao LZ, Clark H, *et al*. Telehealth practice among health centers during the COVID-19 pandemic—United States, July 11–17, 2020. *MMWR Morb Mortal Wkly Rep* 2020; 69 (50): 1902–5.
- DeVoe JE, Likumahuwa-Ackman SM, Angier HE, *et al*. A practice-based research network (PBRN) roadmap for evaluating COVID-19 in community health centers: a report from the OCHIN PBRN. *J Am Board Fam Med* 2020; 33 (5): 774–8.
- Damian AJ, Gonzalez M, Oo M, Anderson D. A national study of community health centers' readiness to address COVID-19. *J Am Board Fam Med* 2021; 34 (Suppl.): S85–94.
- Sudat S EK, Robinson SC, Mudiganti S, Mani A, Pressman AR. Mind the clinical-analytic gap: electronic health records and COVID-19 pandemic response. *J Biomed Inform* 2021; 116: 103715.
- Righi C, James J. *User-Centered Design Stories: Real-World UCD Case Studies*. San Francisco, CA: Elsevier; 2010.
- Health Resources and Services Administration. Uniform Data System (UDS) Resources. 2021. <https://bphc.hrsa.gov/datareporting/reporting/index.html> Accessed April 2, 2021.
- Center for Disease Control and Prevention. Immunization Information Systems (IIS): Core Data Elements for IIS Functional Standards v4.0. 2018. <https://www.cdc.gov/vaccines/programs/iis/core-data-elements/iis-func-stds.html> Accessed April 2, 2021.
- The Office of the National Coordinator for Health Information Technology. United States Core Data for Interoperability (Version 1). 2020. https://www.healthit.gov/isa/sites/isa/files/2020-10/USCDI-Version-1-July-2020-Errata-Final_0.pdf Accessed April 2, 2021.
- Weir RC, Proser M, Jester M, Li V, Hood-Ronick CM, Gurewich D. Collecting social determinants of health data in the clinical setting: findings from national PRAPARE implementation. *J Health Care Poor Underserved* 2020; 31 (2): 1018–35.
- U.S. National Library of Medicine. Value Set Authority Center. 2021. <https://vsac.nlm.nih.gov/> Accessed April 2, 2021.
- The COVID-19 Interoperability Alliance. COVID-19 Resource Downloads. 2021. <https://covid19ia.org/downloads/> Accessed April 2, 2021.
- SNOMED International. SNOMED CT COVID-19 Related Content—Announcements—SNOMED Confluence. 2021. <https://confluence.ihtsdo-tools.org/display/snomed/SNOMED%2BCT%2BCOVID-19%2BRElated%2BContent> Accessed March 18, 2021.
- RxNorm. Product, Program, and Project Descriptions. 2021. <https://www.nlm.nih.gov/research/umls/rxnorm/index.html> Accessed March 18, 2021.
- Centers for Disease Control and Prevention. International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM). 2021. <https://www.cdc.gov/nchs/icd/icd10cm.htm> Accessed April 4, 2021.
- American Medical Association. CPT® (Current Procedural Terminology). 2021. <https://www.ama-assn.org/amaone/cpt-current-procedural-terminology> Accessed March 18, 2021.
- LOINC. SARS-CoV-2 and COVID-19 Related LOINC Terms. 2021. <https://loinc.org/> Accessed March 18, 2021.
- US Department of Health and Human Services. 2015 Edition Health Information Technology (Health IT) Certification Criteria, 2015 Edition Base Electronic Health Record (EHR) Definition, and ONC Health IT Certification Program Modifications. 2015. <https://www.federalregister.gov/documents/2015/10/16/2015-25597/2015-edition-health-information-technology-health-it-certification-criteria-2015-edition-base> Accessed March 18, 2021.
- Center for Disease Control and Prevention. Identify Essential Workers for Public Health Data Collection and Analysis. 2021. <https://www.cdc.gov/niosh/topics/coding/essentialworkers/default.html> Accessed March 18, 2021.
- Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. *JAMA* 2020; 323 (24): 2466–7.
- Kim SJ, Bostwick W. Social vulnerability and racial inequality in COVID-19 deaths in Chicago. *Health Educ Behav* 2020; 47 (4): 509–13.
- Wadhwa RK, Wadhwa P, Gaba P, *et al*. Variation in COVID-19 hospitalizations and deaths across New York City boroughs. *JAMA* 2020; 323 (21): 2192–5.
- The Gravity Project. The Gravity Project: Consensus-Driven Standards on Social Determinants of Health. 2021. <https://confluence.hl7.org/display/GRAV/The+Gravity+Project> Accessed July 9, 2021.
- Centers for Disease Control and Prevention. Adapting Clinical Guidelines for the Digital Age. 2020. <https://www.cdc.gov/dphss/clinical-guidelines/index.html> Accessed July 9, 2021.