Title
Capabilities and consequences of data mapping in emergent health scenarios: Using a multi-site COVID-19 research data set as an example

Author
Kothari, Shikha Yashwant

Publication Date
2023

Peer reviewed|Thesis/dissertation

Capabilities and consequences of data mapping in emergent health scenarios: Using a multi-site COVID-19 research data set as an example

By

SHIKHA YASHWANT KOTHARI

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Health Informatics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

JAMES MARCIN, Chair

PRABHU SHANKAR

MARK CARROLL

Committee in Charge

2023

TABLE OF CONTENTS

TABLE OF TABLES

ABSTRACT

During the Coronavirus Disease 2019 (COVID-19) pandemic, a public health emergency (PHE) was declared by the United States (U.S.) government, reducing the number of in-person clinic visits and increasing telemedicine utilization.[1–12] Healthcare reimbursement guidelines evolved on an ongoing basis and a lack of standardization in procedure coding for telemedicine visits created confusion amongst providers.[13–17]

This thesis focuses on a standardized, multi-site data repository, the University of California (UC) COVID-19 Research Dataset (UC CORDS) and uses it as an example to review the downstream consequences of ad-hoc data mapping of new services such as telemedicine visits to formalized coding systems during the COVID-19 pandemic. The findings are then translated to recommendations for creating best practices to combat challenges associated with building computable phenotypes for complex multi-site data in emergent health scenarios.

Included patients had a COVID-19 test result mapping to the designated LOINC codes between Feb 2020 to Feb 2021. My study results reflect the lack of standardization in standard vocabulary naming conventions and concept mapping for telehealth. This makes it difficult for researchers to find telehealth-specific data from CDM datasets like UC CORDS, which only capture data mapped to standard vocabularies. My journey through this master's thesis also highlights the multiple data access, data fluency, and data management challenges that clinical researchers face with complex healthcare datasets such as UC CORDS.

In conclusion, although telemedicine has been considered beneficial for several years, the COVID-19 pandemic offered the best opportunity to improve telemedicine services and fully integrate them into healthcare reimbursement workflows and healthcare information systems. Based on the outcomes of this study, there is still room for process improvement in regard to handling the needs of data capture for new services in emergency scenarios, and healthcare institutions should involve multiple key stakeholders at an earlier stage when developing and implementing a digital infrastructure.

BACKGROUND

Explosive growth of Telemedicine during the Coronavirus Disease 2019 Pandemic

In early 2020 as the Coronavirus Disease 2019 (COVID-19) pandemic grew in size, telemedicine grew astronomically.[17,38–40] This pandemic was different compared to previous outbreaks, partly due to the expanded reach of public health information and news regarding COVID-19 fatalities through social media platforms. Increased understanding of the COVID-19's morbidity and mortality led to the rapid implementation of social distancing and lock-down measures by local, state and national governments worldwide. Life around the world went online, especially school and work. However, a majority of healthcare institutions were not online, and nor were healthcare providers. Hospitals and clinics found themselves overflowing with patients and reported record numbers of COVID-19 cases. Nosocomial spread of COVID-19 to previously healthy patients catapulted stringent infection control measures and the cancelling of elective procedures and any unnecessary outpatient visits. This led to an unprecedented rise in telemedicine utilization, especially for services that did not need an in-person visit.

Effect of the Coronavirus Disease 2019 on Telemedicine Reimbursement

Telemedicine Reimbursement during the Coronavirus Disease 2019 Pandemic

Medicare Population at Risk

As time progressed, data on the effects of COVID-19 on Medicare patient morbidity and mortality became available. Given the unknown nature of the COVID-19 virus and lack of available data regarding its risk factors, disease manifestations and outcomes, many clinical studies were carried out and research findings disseminated on an ongoing basis through online platforms. Elderly patients, immunocompromised patients, patients with underlying medical conditions and patients at skilled nursing facilities (SNFs) were found to be at high risk for severe complications and death from COVID-19. Elderly adults are pre-disposed to poor immunity and multiple comorbidities which place them at high risk for severe complications and/or death from COVID-19. Elderly adults with chronic medical conditions are also the patients most likely to require long-term medical care and inhabit SNFs and assisted living facilities, which are all high-risk factors for contracting COVID-19 as well. Given that all elderly patients above 65 years of age are covered by Medicare, the Medicare population as a whole classified as high-risk. With the patient home not an eligible originating site under Medicare and with healthcare facilities seen as high-risk locations, elderly patients struggled to receive care. In addition, practices struggled financially with the reduction in in-person visits and lack of telehealth reimbursement.[41]

Declaration of a Public Health Emergency and the Center for Medicare and Medicaid Services 1135 Waiver

Because of the rapid dissemination of such findings, CMS issued the 1135 waiver in March 2020. In light of the federal PHE, the 1135 waiver (1) expanded Medicare coverage to include the patient home as an eligible originating site for telehealth services, (2) created new codes specifically for telemedicine encounters which removed the requirement for GT/GQ/G0/95 modifiers and the POS 02 code, and (3) increased the reimbursement payment for telemedicine encounters so that it would equal the reimbursement payment for in-person encounters. State Medicaid programs and commercial payors also followed suit by incorporating similar coverage policies. This led to increased incentive for practices to provide telemedicine services, as well as increased investment towards the improvement of telehealth technology, equipment and infrastructure at a nationwide level.

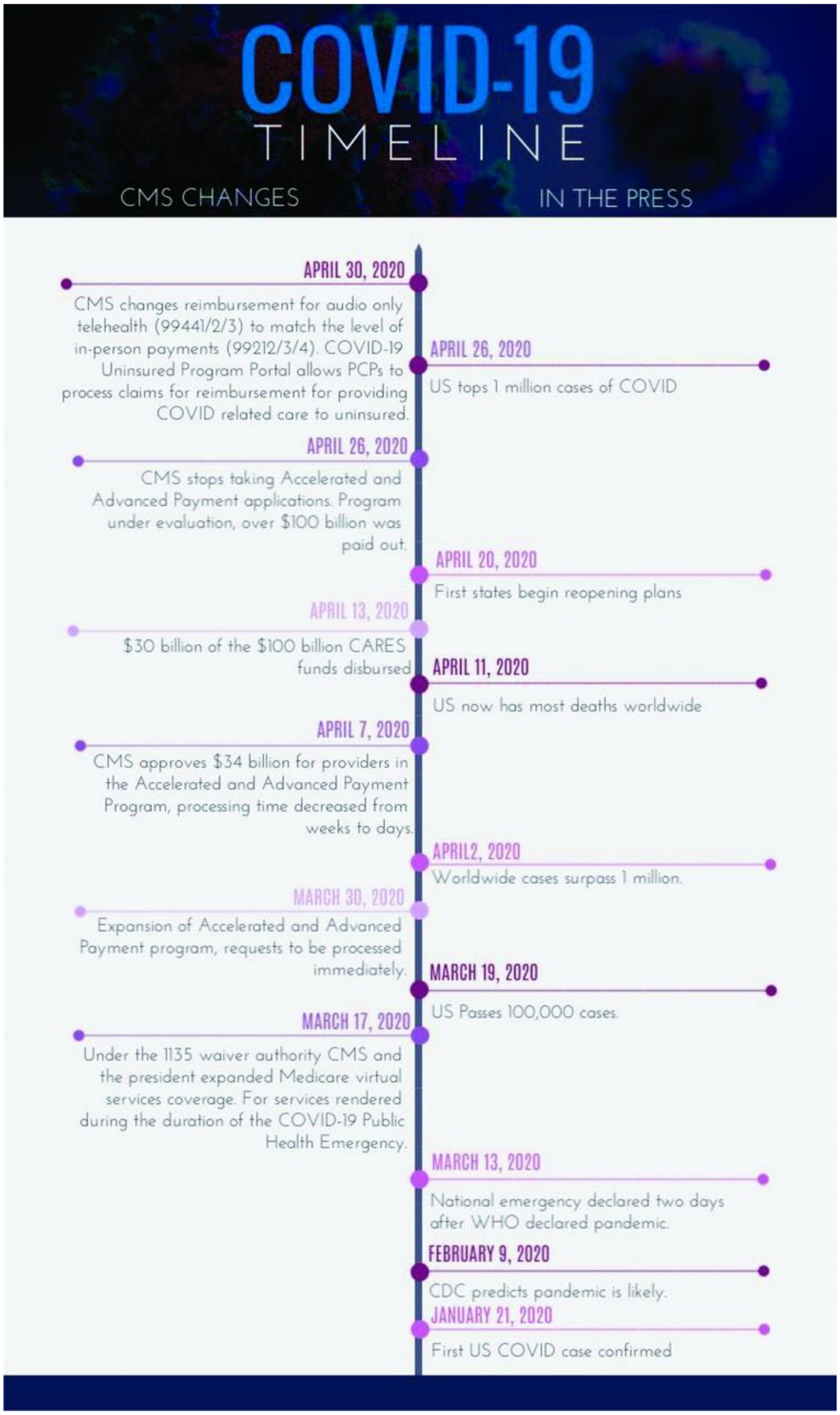

Figure 1 "A descriptive timeline of the changes made by Centers for Medicare and Medicaid Services (CMS) during the COVID-19 pandemic juxtaposed against popular events in the press at the time."[41]

Effect of the Coronavirus Disease 2019 on Research

Centralized Data Initiatives

COVID-19 pandemic accelerated many ongoing data efforts, including an initiative by the University of California (UC) Office of the President (UCOP) to standardize and aggregate healthcare data from the six UC health systems into a single UC Health Data Warehouse (UCHDW). The UCHDW initiative was already ongoing pre-pandemic, but efforts accelerated during the COVID-19 pandemic as the importance of health information exchange (HIE) became more apparent. Similarly, researchers and healthcare institutions across the nation collaborated on centralized data initiatives such as the National COVID Cohort Collaborative (N3C) to collect, deidentify and aggregate massive amounts of secure healthcare data for covid-19 related research.[42]

National Coronavirus 2019 Cohort Collaborative

The N3C is a partnership of multiple programs supported by the National Center for Advancing Translational Sciences (NCATS) and the National Institute for General Medical Sciences (NIGMS), under the National Institutes of Health (NIH).[43–45] The N3C Data Enclave is a central, harmonized collection of deidentified patient data electronically retrieved from the EHR's of several participating healthcare institutions in order to aid COVID-19 research efforts. It is the largest collection of COVID-19 data and follows the Observational Medical Outcomes Partnership (OMOP) common data model (CDM).[43–45]

University of California Health Data Warehouse and related initiatives

Coordination of Healthcare Data at Scale: Development of a Multi-Institutional Data Repository

The planning and development of UCHDW was a massive effort by dedicated teams at each UC Health site and occurred over the span of several years starting in 2019, once all six included health systems had completed implementation of Epic as the primary EHR software.[46–48] The implementation of the same health information system at each health system allowed for a common data architecture for each site, facilitating data mapping and reducing the amount of pre-processing required for the aggregation of multi-institutional healthcare data. The COVID-19 pandemic accelerated these efforts due to increased awareness that massive data repositories would be required in order to carry out large-scale research studies with higher accuracy.[46,48] The development of N3C was an additional contributing factor since institutions participating in N3C were required to convert their unstructured EHR data to structured data collections compatible with the OMOP CDM.

University of California Coronavirus 2019 Research Data Set

Because the UCHDW data is identifiable and contains protected health information (PHI), it is housed on secure servers with multiple safeguards and highly controlled access. Access requires institutional ethics approval as well as a qualified ETL analyst to extract the data for researchers in a secure Azure environment called Data Bricks. In 2020, in light of the amount of time that the conventional extract-transform-load (ETL) process can take and considering the pandemic and its time-sensitive nature, data scientists and informaticists at the UC Office of the President (UCOP) decided to create a more accessible dataset to promote and accelerate COVID-19 related research.[46] A subset of UCHDW data on patients who were tested for COVID-19 was extracted and statistically deidentified in accordance with the HIPAA Safe Harbor Law, with the exception of dates which are left unmasked, forming an LDS known as the University of California Coronavirus 2019 Research Data Set (UC CORDS).

A limited data set (LDS) is defined by the HIPAA Privacy Rule as PHI that excludes certain direct identifiers of an individual or of relatives, employers or household members of the individual — but may include city, state, ZIP code and/or elements of dates. To understand this more, there are eighteen unique identifiers considered by the HIPAA Privacy Rule to be PHI: 1) names, 2) geographic data, 3) all elements of dates, 4) telephone numbers, 5) FAX numbers, 6) email addresses, 7) Social Security numbers (SSN), 8) medical record numbers (MRN), 9) health plan beneficiary numbers, 10) account numbers, 11) certificate/license numbers, 12) vehicle identifiers and serial numbers including license places, 13) device identifiers and serial numbers, 14) web URLs, 15) internet protocol addresses, 16) biometric identifiers (i.e. retinal scan, fingerprints), 17) full face photos and comparable images, and 18) any unique identifying number, characteristic or code. An LDS must be devoid of at least seventeen unique identifiers classified as PHI to be considered in compliance with the HIPAA Privacy Rule. Additionally, an LDS can be disclosed only for purposes of research, public health or health care operations. One advantage of utilizing an LDS is that an institutional ethics review is not mandatory. This reduces barriers for initiating and carrying out a research study, which in turn allows for a timelier publication of findings.

Telemedicine Implementation, Telemedicine Data Quality, and Telemedicine Research

With the accelerated growth of telehealth during the pandemic, many groups were interested in telemedicine related data for research purposes. However the capture and storage of telemedicine data was not standardized. This is in part due to confusion among healthcare providers as well as billers and coders on how to code telemedicine services and the slow development and integration of digital infrastructures to support telemedicine related data capture in EHR’s.

Adoption of and adherence to new guidelines is key to effective implementation of changes. The CMS 1135 waiver was beneficial for practices well-equipped for telemedicine, however for practices that were ill-equipped it took time for new billing and coding workflows to be developed and implemented. With telemedicine coding already as complex as it was pre-pandemic, the new changes led to confusion amongst healthcare providers and a lack of uniformity in telemedicine coding practices among healthcare institutions.

It also took time for the development of a digital infrastructure to support the new workflows. Once a coding system is validated for a new service, it can take anywhere from months to years for the development and integration of data fields in each EHR system to support the additional code(s) and/or coding system. While this transition process occurs, a temporary method, or stop gap is utilized to capture relevant information until the infrastructure is in place for the new system.

In the case of the COVID-19 pandemic, the new service was telemedicine and the CMS 1135 waiver presented new codes for telemedicine visits as well as an updated reimbursement system. Until computable phenotypes and EHR infrastructure were developed, validated and implemented the temporary stop gap at many healthcare practices was to code the in-person equivalent of each E/M service performed via telemedicine with relevant modifiers and/or POS codes and request a reimbursement amount equal to the amount reimbursed for the in-person service. This became a coding standard for several months. As computable phenotypes were created, the digital infrastructure of various EHR’s adapted to incorporate these phenotypes and corresponding front-facing data fields for the new CMS telemedicine codes. However, providers found it overwhelming to adapt to the continuously evolving changes while simultaneously focusing on patient care and rising COVID-19 cases.

With the EHR infrastructure at each healthcare institution in different stages of development, validation and/or implementation, along with various provider coding workflows, telemedicine data quality in EHR's suffered. The non-standardized capture of telemedicine visit-related data presented a challenge for clinical researchers attempting to determine a standardized methodology for ascertaining which visits involved telemedicine.

Context/Significance

My initial research question was related to demographic and socioeconomic differences in the patient population for telehealth encounters versus in-person encounters captured by the UC CORDS LDS. However, while I could find data on office visits and inpatient hospital admissions, I could not find any visit occurrences that mapped to visit subtype concepts. I then explored the online UCDHW Data Documentation available to researchers accessing UC CORDS, which contained helpful interactive tools created in Tableau such as a Code Mapping Explorer for mapped and unmapped concepts in UC CORDS.[49] I noted that in the Code Mapping Explorer there are standardized codes for telemedicine visits which are mapped in the UCDH OMOP, however they are mapped to visit subtype rather than visit type.

Visit subtype is only available in an extension table of the identified OMOP, and extension tables were excluded from UC CORDS. Extension tables reflect data types which have not been harmonized in a standardized fashion by the OHDSI OMOP community.[46,50] Such tables were included in the identified OMOP for operational purposes; however, since the data types were not standardized, they were excluded from UC CORDS.[46] To obtain visit subtype information, I was told I would need to obtain IRB approval and wait for a qualified ETL analyst from the identified OMOP team to query the identified OMOP and extract the data I required.

Given the complexity and length of the proposed timeline on accomplishing this, I decided not to move forward with my original research question and instead focused on evaluating the consequences and capabilities of UC CORDS data mapping that emerged during the COVID-19 pandemic. My study proposes desiderata for best practices in combating challenges associated with building computable phenotypes for complex multisite data.

AIMS/OBJECTIVES

Objectives

1. Characterize the distribution of ad-hoc codes and mapped standardized codes from formal coding systems once developed and released.
2. Use the example of UC CORDS and telehealth coding to review the downstream consequences of this mapping and make recommendations for future phenotyping of new conditions in multi-site data repositories.

METHODS

Study Type

Cross-sectional study design with descriptive statistics regarding distribution of standard concepts for procedure occurrence representations across concept classes and domains.

Time Period

February 12, 2020 (date of first COVID-19 case at a UC Health site - UC Davis Health) to February 11, 2021.

Data Source
The data source for this study is the UC COVID-19 Research Data Set (UC CORDS), Limited Data Set (LDS). UC CORDS LDS is a relational database based on the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

Study Sites
The data in UC CORDS is collected, aggregated and published by dedicated informatics and data science teams at the University of California Health System (UC Health), which is comprised of 19 health professional schools, five academic medical centers and 12 hospitals across the state of California.

The academic medical centers contributing data to UC CORDS are University of California, San Diego Health System (UCSD Health), University of California, Davis Health System (UC Davis Health), University of California, Los Angeles Health System (UCLA Health), University of California, San Francisco Health System (UCSF Health), and University of California, Irvine Health System (UC Irvine Health). Electronic medical records (EMR) from Epic for patients tested for COVID-19 at each of these academic medical centers, as well as their affiliated UC Health care sites, are the source of this data.

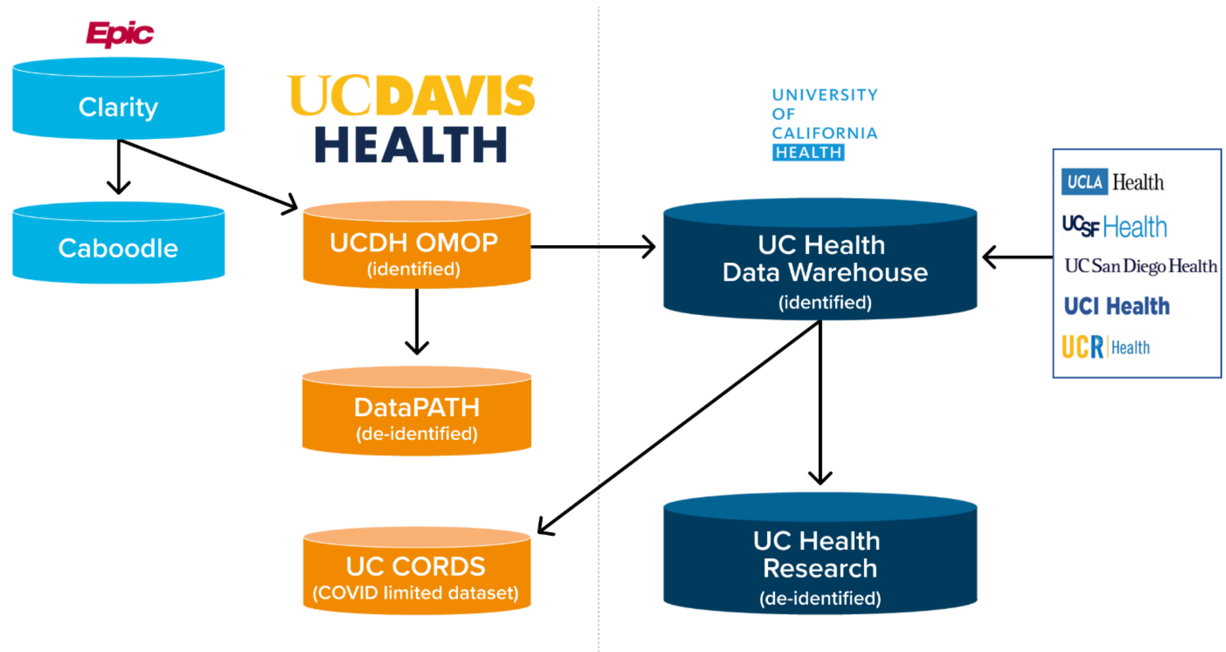

Figure 2 Visual representation and data pipeline for the various sources of clinical data from Epic at UC Davis Health as well as the entire University of California Health.[51]

Detailed Data Source Description and Pipeline
Data is retrieved from the Epic electronic medical record at each UC health site and programmatically processed and loaded into a single data warehouse at the UC Health level, known as the UC Health Data Warehouse (UCHDW). UCHDW is refreshed on a monthly basis. UC CORDS data is programmatically generated from UCHDW on a weekly basis.

Raw Electronic Medical Records Data Capture, Storage, Retrieval and Pre-Processing
At each participating health system, patient data is captured in real-time by the Epic EMR. Initially this real-time data is housed in a transactional database called Chronicles. Data from Chronicles is

programmatically processed through Extract – Transform – Load (ETL) and loaded into an analytical database called Clarity at a specific time each evening. Clarity contains historical data from a single institution's Epic EMR. Once the Chronicles to Clarity ETL is complete, data from Clarity is also programmatically processed through ETL each night. Once transformed, the Clarity data is loaded into a data warehouse which is also institution specific and known as Caboodle. Apart from Clarity data, Caboodle also contains data from other sources, such as hospital registries and linked patient data from outside hospitals which utilize Epic. The data architecture of Caboodle is similar to Clarity.

Transformation of Pre-Processed Data and Loading a Common Data Model Database

Aside from the daily export of Clarity data into Caboodle, raw data from Clarity at each institution is also loaded into another local relational database following the OMOP CDM. For example, at UC Davis Health this database is known as the UC Davis Health (UCDH) Identified OMOP. The OMOP CDM follows a relational database model with Person (patients) as the observational unit and the records in each table mapped to standard concepts. Before loading it into UCDH Identified OMOP, the raw data from Clarity at UC Davis Health is transformed and mapped to various tables of the OMOP CDM. The local identified OMOP data from each participating UC Health site is loaded into the UCDHW on a monthly basis.

University of California Coronavirus Disease 2019 Research Data Set

UC CORDS LDS is programmatically generated from the UCHDW.

Inclusion Criteria

Data Source Inclusion Criteria

The information provided in Table 1 has been taken from the UC CORDS public documentation[47]

Table 1 UC CORDS LDS Inclusion Criteria

"**UC CORDS Description**

• A summary of daily SARS-CoV-2 testing, results, and hospitalization data from each site

• All available historical clinical data associated with these patients from the monthly UCHDW refreshes

• Only data mapped to a standard terminology is available.

• All identity key fields are not persistent and will be regenerated each week

Patient Inclusion Criteria

All patients that tested positive or negative for COVID-19 are included in the dataset from the monthly daily data feed (OMOP table: person). Patient demographics and other patient attributes are stored in the person table. Restricted patients identified by sites are removed from the dataset.

SARS-CoV-2 Testing

- PCR tests from monthly files are mapped to one of the following LOINC codes: 94309-2, 94500-6, 94531- 1, 94306-8, 94534-5, 94559-2, 94533-7

Note: Only data mapped to a standard terminology is available in UC CORDS. OMOP CDM represents each unique data value as a concept which is linked to a standard vocabulary such as a terminology or ontology. Example: patient with hypertension as a diagnosis in their medical record, the hypertension can be represented as an ICD-10 Diagnosis code (terminology), or as a clinical observation code SNOMED (hierarchical, ontology). These are standard vocabularies."

Study Inclusion Criteria

The subset of data included in my study is the output of a SQL join involving two tables: the procedure_occurrence table in UC CORDS joined with the concept table in UC CORDS. Only OMOP concepts mapping to a procedure_occurrence_concept_id in the UC CORDS procedure_occurrence table were reviewed.

Institutional Ethics approval

Institutional Review Board (IRB) review was not required for this study. Access to this dataset is limited to staff and researchers affiliated to a UC Health campus. A data use agreement must be signed before accessing the dataset; however, given that it is an LDS, no institutional review board (IRB) approval is required for access to the UC CORDS data. Researchers seeking to access the dataset must be on a secure UC-provided network at a UC health system campus, or a UC-provided virtual private network (VPN).

Data Extraction

Accessing Dataset

I obtained access to UC CORDS by submitting a request ticket which included the following required information: my purpose for accessing the data, professional title, years of experience with data, preferred environment to access the data, and consent to sign a data use agreement before being provisioned access.

Environment

Environments I was provided for accessing the data: (1) Windows 10 Virtual Machine (VM) accessed through connecting with remote desktop and (2) JupyterHub built on Linux platform, accessed from a browser via a secured https URL.

Dataset Exploration

To explore data from UC CORDS, researchers must retrieve data with structured query language (SQL). The dataset is a database is housed on Microsoft SQL Server, which can be directly explored and queried in the MS SQL Server Management Studio software while remotely connected to the secure VM.

Researchers can also query the database using the Python programming language in the secure browser-based JupyterHub, or with other platforms in the secure VM such as R programming language using RStudio software. Both of these are accomplished with the help of an application programming interface (API). A combination of Python and R APIs, as well as direct SQL queries with MS SQL Server Management Studio, were utilized for this study.

Overview of My Approach

I connected to the database using secure credentials provided by the UC CORDS team.

I initially utilized the secure JupyterHub environment to query and explore the database with Python programming language. Querying was possible with the help of a package created by the UC Davis UC CORDS team, which served as an API to connect to and query the database.

In the VM I utilized MS SQL Server Management Studio for direct querying of the data with SQL, and utilized RStudio to connect to and query the database with R programming language. MS SQL Server Management Studio was used for a more interactive representation of the database, as well as to test the outputs of SQL queries before inserting them into R and python code.

The final results, figures and observations in this thesis were generated through R programming with RStudio.

Mapping Procedure Occurrence Concepts

I worked with the UC CORDS team and drafted a SQL query which joined the following tables of UC CORDS:

Tables Joined

The following UC CORDS tables were joined: procedure_occurrence and concept.

The column names, or attributes, for each table, or entity, are in Table 2 and Table 3.

Table 2: List of Attributes in the procedure_occurrence table of UC CORDS LDS

Attributes: procedure_occurrence Table
procedure_occurrence_id
person_id
procedure_concept_id
procedure_date
procedure_datetime
procedure_type_concept_id
modifier_concept_id
quantity
provider_id
visit_occurrence_id
visit_detail_id
procedure_source_value
procedure_source_concept_id
modifier_source_value

Table 3: List of Attributes in the concept table of UC CORDS LDS

Attributes: concept Table
concept_id
concept_name
domain_id
vocabulary_id
standard_concept
concept_code
valid_start_date
valid_end_date
invalid_reason
concept_class_id

Construction of Join Query

The SQL join was accomplished by joining procedure_occurrence to the concept table twice.

Keys

Keys from each table were utilized to link the two tables. Primary key was concept_id from the concept table. This served as a foreign key to the procedure_occurrence table, linking to both the procedure_concept_id and procedure_type_concept_id .

Join Components

(1) First join to map each procedure_concept_id to its corresponding concept_id in the concept table.
   a. Joined procedure_occurrence.procedure_concept_id = concept.concept_id
   b. This mapped all of the procedure occurrences to their corresponding concept.

(2) Second join to map each procedure_type_concept_id to its corresponding concept_id in the concept table.
   a. Joined procedure_occurrence.procedure_type_concept_id = concept.concept_id
   b. This mapped all of the procedure occurrence types to their corresponding concept.

Query Output

Table 4 demonstrates the first two rows, or head, of the final output table. This output table represents the final dataset explored with R programming to perform descriptive statistics for this study.

Table 4: Head of the final SQL query output

OMOP_TABLE	CONCEPT_SOURCE	domain_id	vocabulary_id	concept_class_id	concept_name	concept_code	concept_id	standard_concept	Total_Occurrences
PROCEDURE_OCCURRENCE	Hospitalization Cost Record	Measurement	CPT4	CPT4	Zinc	84630	2212629	S	9603
PROCEDURE_OCCURRENCE	Hospitalization Cost Record	Procedure	CPT4	CPT4	Dual-energy X-ray absorptiometry (DXA), bone density study, 1 or more sites; axial skeleton (eg, hips, pelvis, spine)	77080	2211826	S	13310

Each concept_id maps to a unique procedure; however the same procedure_concept_id can appear multiple times in this dataset if the source of the procedure data is different (CONCEPT_SOURCE), for example "Hospitalization Cost Record", "Primary Procedure", "Referral Record", "EHR order list entry", "Procedure recorded as diagnostic code", "Flowsheet Procedure", "Health Maintenance", etc.

Standard Concepts Column

According to the UCOP Data Dictionary for the UCHDW / UC CORDS dataset, in the standard_concept column of an OMOP concept table, there are three possible values: "S" (Standard Concept), "C" (Classification Concept), or "NA" (Non-Standard Concept).

From the Book of OHDSI https://ohdsi.github.io/TheBookOfOhdsi/StandardizedVocabularies.html# , which describes the OMOP Common Data Model in detail, this is the exact, word-for-word description provided regarding Standard, Non-Standard, and Classification concepts:

5.2.6 Standard Concepts

One concept representing the meaning of each clinical event is designated the Standard. For example, MESH code D001281, CIEL code 148203, SNOMED code 49436004, ICD9CM code 427.31 and Read code G573000 all define "Atrial fibrillation" in the condition domain, but only the SNOMED concept is Standard and represents the condition in the data. The others are designated non-standard or source concepts and mapped to the Standard ones. Standard Concepts are indicated through an "S" in the STANDARD_CONCEPT field. And only these Standard Concepts are used to record data in the CDM fields ending in "_CONCEPT_ID".

5.2.7 Non-Standard Concepts
Non-standard concepts are not used to represent the clinical events, but they are still part of the Standardized Vocabularies, and are often found in the source data. For that reason, they are also called "source concepts". The conversion of source concepts to Standard Concepts is a process called "mapping"… Non-standard concepts have no value (NULL) In the STANDARD_CONCEPT field.

5.2.8 Classification Concepts
These concepts are not Standard, and hence cannot be used to represent the data. But they are participating in the hierarchy with the Standard Concepts, and can therefore be used to perform hierarchical queries.

Final Data Extraction and Visualization with R

Data Extraction

An R package dbmi() which served as an API to MS SQL Server databases was utilized to connect to the UC CORDS database. This R code was contained in a separate script with the secure credentials. db_getquery() was populated with the finalized SQL query string and the final output dataset was obtained.

Data Visualization

Packages tidyverse() and dplyr() were used to processs and slice the dataset. ggplot2() was utilized for creating plots.

One analysis I performed was an empirical exploration of which concepts in the procedure_occurreence table have the word "telehealth" somewhere in the concept_name value.

RESULTS

The results presented in this study characterize the distribution of OMOP standard concepts available for mapped procedure representations in the procedure_occurrence table of the UC CORDS LDS dataset, across standard vocabularies, concept classes and concept domains. They also identify which telemedicine and telehealth related OMOP standard concepts, standardized vocabularies and codes were documented for procedure occurrences in the Epic EHR across the UC Health systems between February 2020 to February 2021. Included patients had a COVID-19 test result mapping to the designated LOINC codes between Feb 2020 to Feb 2021.

The final query output utilized for data visualization was a 10 column data frame with 49,895 rows. These rows represent the unique procedure concept and procedure type concept representations available in the procedure_occurrence table of UC CORDS, with the total number of occurrences for each representation available as an aggregate sum.

Primary Outcomes

1. Which standard vocabularies, concept classes and concept domains are available for mapped procedure occurrence concepts in the OMOP concept table of the UC CORDS LDS?

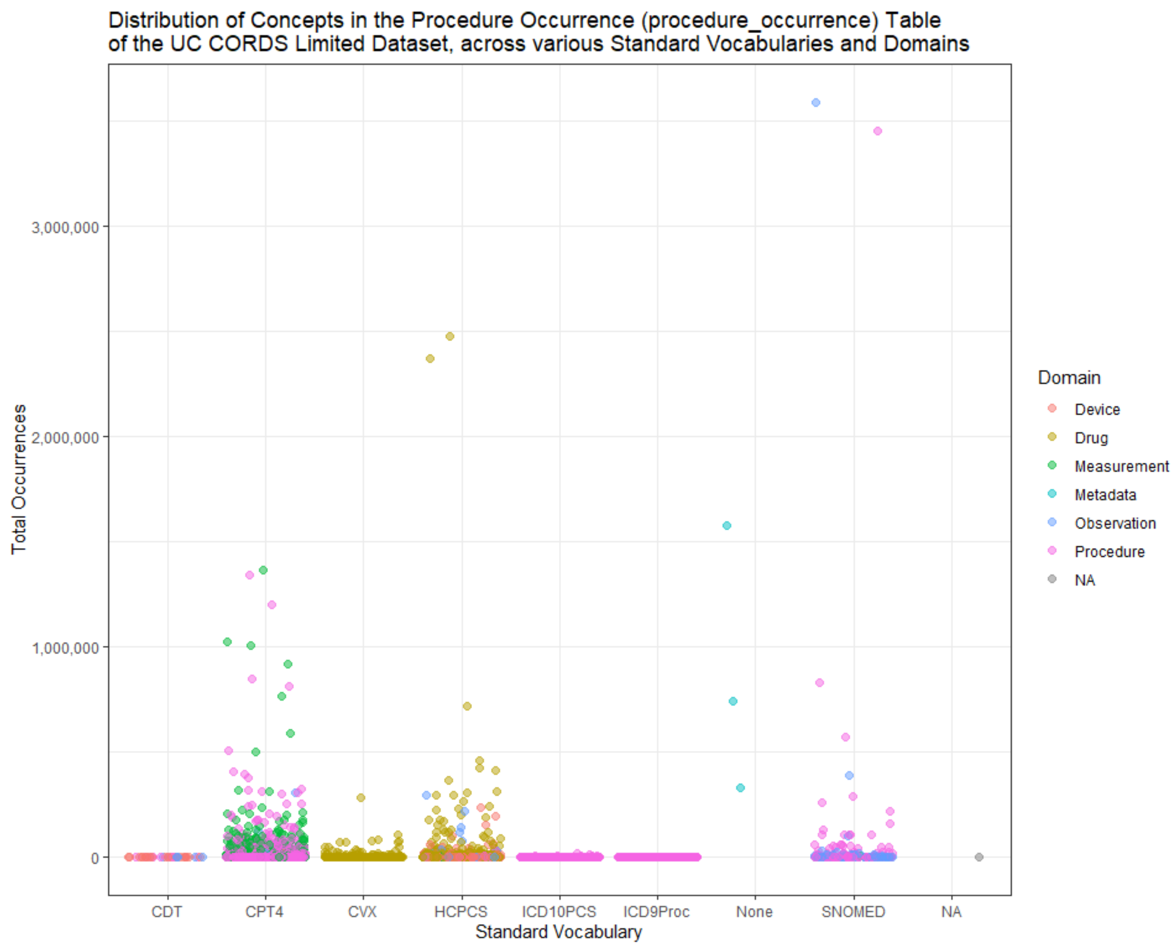

Figure 3 Distribution of Concepts in the Procedure Occurence (procedure_occurrence) Table of the UC CORDS Limited Dataset, across various Standard Vocabularies aned Domains. CDT, Code on Dental Procedures and Nomenclature. CPT4, Current Procedure Terminology 4th Edition. CVX, HL7 Table 0292, Vaccine Administered. HCPCS, Health Care Common Procedure Coding System. ICD10PCS, International Classification of Diseases, 10th Revision: Procedure Coding System. ICD9Proc, International Classification of Diseases, 9th Revision. SNOMED, Systemized Nomenclature of Medicine Clinical Terms.

2. Within each standard vocabulary available for mapped procedure occurrence concepts in the OMOP concept table of the UC CORDS LDS, what are the unique concept classes and unique concept domains available? What is the distribution of concepts across these concept classes and concept domains?

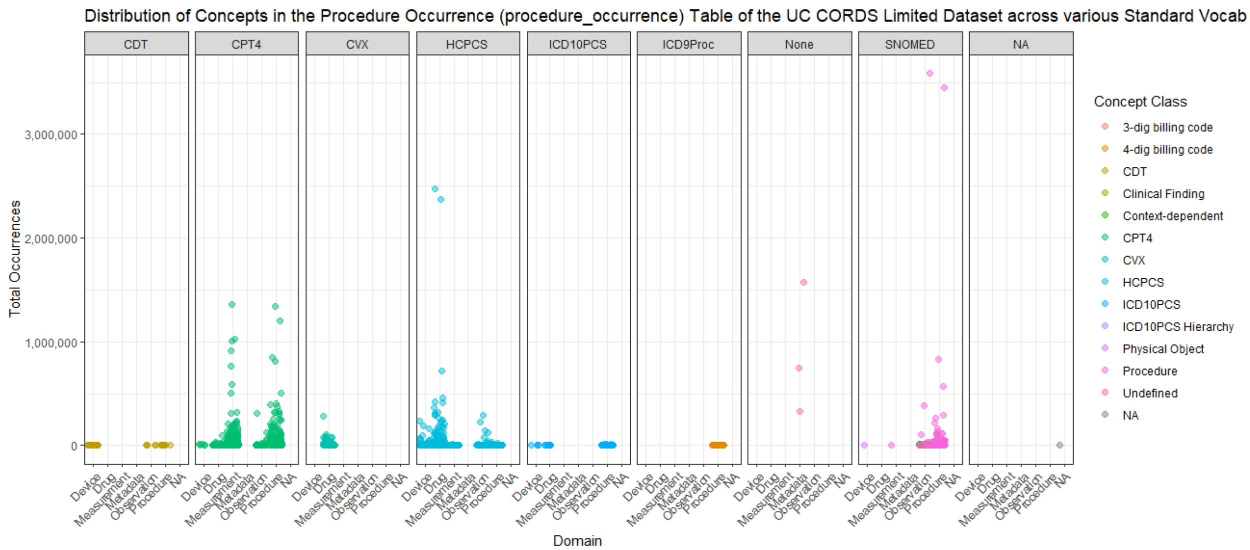

Figure 4 Distribution of Concepts in the Procedure Occurrence (procedure_occurrence) Table of the UC CORDS Limited Dataset, across various Standard Vocabularies, Domains and Concept Classes. CDT, Code on Dental Procedures and Nomenclature. CPT4, Current Procedure Terminology 4th Edition. CVX, HL7 Table 0292, Vaccine Administered. HCPCS, Health Care Common Procedure Coding System. ICD10PCS, International Classification of Diseases, 10th Revision: Procedure Coding System. ICD9Proc, International Classification of Diseases, 9th Revision. SNOMED, Systemized Nomenclature of Medicine Clinical Terms.

3. What are the unique concepts containing the string “telehealth” in the procedure occurrence table of UC CORDS? What are the unique concept domains for these concepts? Within each concept domain, what is the total number of occurrences for each concept and what are the standardized vocabularies and codes mapped to each concept?

List of unique domains for procedure occurrence concepts in the procedure_occurrence table containing the string “telehealth” in the procedure concept name

Table 5: List of unique domains for procedure occurrence concepts in the procedure_occurrence table containing the string “telehealth” in the procedure concept name

Unique Concept Domains of Procedure Occurrence Concepts Containing the String “telehealth” in the Concept Name
Procedure
Observation

List of unique standard vocabularies for procedure occurrence concepts in the procedure_occurrence table containing the string “telehealth” in the procedure concept name

Table 6: List of unique standard vocabularies for procedure occurrence concepts in the procedure_occurrence table containing the string “telehealth” in the procedure concept name

Unique Standard Vocabularies for Procedure Occurrence Concepts Containing the String “telehealth” in the Concept Name
HCPCS

Total occurrences of each unique telehealth procedure representation in the procedure_ocurrence dataset when filtered for a concept domain = Procedure

Table 7: Total occurrences of each unique telehealth procedure representation in the procedure_ocurrence dataset when filtered for a concept domain = Procedure

Unique Procedure Occurrence Concepts Containing the String "telehealth" in the Concept Name and Having a Concept Domain of "Procedure"	Total Occurrences
Telehealth consultation, emergency department or initial inpatient, typically 50 minutes communicating with the patient via telehealth	10
Telehealth consultation, emergency department or initial inpatient, typically 30 minutes communicating with the patient via telehealth	14
Telehealth consultation, critical care, initial , physicians typically spend 60 minutes communicating with the patient and providers via telehealth	1
Telehealth consultation, emergency department or initial inpatient, typically 70 minutes or more communicating with the patient via telehealth	6

Total occurrences of each telehealth procedure representation in the procedure_ocurrrence dataset when filtered for concept domain = Observation

Table 8: Total occurrences of each telehealth procedure representation in the procedure_ocurrrence dataset when filtered for concept domain = Observation

Unique Procedure Occurrence Concepts Containing the String "telehealth" in the Concept Name and Having a Concept Domain of "Observation"	Total Occurrences
Follow-up inpatient consultation, intermediate, physicians typically spend 25 minutes communicating with the patient via telehealth	173
Follow-up inpatient consultation, limited, physicians typically spend 15 minutes communicating with the patient via telehealth	397
Follow-up inpatient consultation, complex, physicians typically spend 35 minutes communicating with the patient via telehealth	46

Secondary Analysis of Findings & Limitations

As mentioned earlier in Table 1, UC Health included existing EHR historical data dating back to 2012 for all included patients in UC CORDS. Of note, as of February 11, 2021 a decision was made by UC Health to limit the historical data for COVID-19 negative patients in UC CORDS to 2019 onwards instead of all past historical data, in an effort to reduce the growing database size. Therefore, the procedure occurrence data in this study likely has higher representation from COVID-19 positive patients. I did not include a specific date range for the procedure occurrences when building my SQL query, so the aggregate values for each procedure occurrence in the query outcome include data on procedures which occurred pre-pandemic. However, I did not consider this to be a limitation since this study's aim was to review data capture, storage and distribution at a specific point in time, not the date/time each procedure occurred.

A prominent limitation of the study is concept categorization and mapping issues. To interpret the study results, one must look at figure 1 again for the timeline of events occurring during the COVID-19 pandemic and the corresponding timeline of CMS policy changes. Early in the pandemic and at the time of data extraction, the CMS 1135 waiver designated approved codes for audio/visual and audio-only synchronous telehealth consultations but also allowed for equal reimbursement of these services regardless of the mode of delivery being in-person or through telehealth. This served as a temporary stop gap to allow time for healthcare institutions to incorporate the new coding recommendations into their EHR infrastructure and clinical workflows. For healthcare facilities and practices which did not yet incorporate the new telehealth codes and recommended workflows, the equal reimbursement allowed providers to bill the designated “Office or other outpatient visit” E/M codes for telemedicine visits without the need for modifiers and claim the same reimbursement amount as the equivalent in-person service. Because of this, providers coded telehealth visits as office visits, so the total procedure code counts for procedure representations with “telehealth” in the concept name do not accurately reflect how many of each type of telehealth encounters actually occurred - the actual number may be higher.

To highlight this further, in Table 6 HCPCS is the only standard vocabulary for all procedure_occurrence concepts with the string “telehealth” in the concept name. As demonstrated in Table 7 and Table 8, the concept names for the various procedure_occurrence concepts with a string match for “telehealth” all mention either “emergency department,” “inpatient,” or “critical care” which implies that these are all hospital related telehealth consultations. Given that none of the concepts containing the string “telehealth” appear to be outpatient visits, it is likely that outpatient services were billed utilizing the CPT4 codes designated by the CMS for “Office or other outpatient visit” E/M services provided by synchronous, face-to-face telehealth. The concept names for these codes do not contain the string “telehealth” and were not incorporated into my limited string search criteria.

Additionally, before developing my string search criteria I reviewed generated a list of the unique concept names for concepts in the UC CORDS procedure_occurrence table and found several string matches for “telehealth”, but did not find any string matches for “telemedicine” in procedure_occurrence table concepts. That being said, concepts with a concept name containing the string “telemedicine” do exist in the UC CORDS concept table; these could be either unmapped or mapped to concepts in tables other than procedure_occurrence. Similarly, OMOP concepts containing “telehealth” in the concept names may span multiple domains and tables across the database, rather than just in the procedure occurrence table as CPT4 / HCPCS codes. Therefore, it is likely that the query I pulled from the procedure occurrence table alone is not sufficient to characterize telehealth-related data capture in UC CORDS at the time of data extraction.

Figure 3 demonstrates that the procedure_occurrence table in UC CORDS is comprised of concepts across multiple domains and multiple standard vocabularies. According to the Book of OHDSI which is the principal resource regarding the OMOP CDM, a concept domain indicates which table the concept is found in. It appears that aside from the “Procedure” domain concepts, there are “Device”, “Drug”, “Measurement”, “Metadata”, “Observation”, as well as null domain concepts in the procedure_occurrence table of UC CORDS as well.

It appears that the few discrete concepts mapped to the “None” standard vocabulary all map to the Metadata domain, each with hundreds of thousands to over a million total occurrences. The Book of OHDSI defines metadata as “A set of data that describes and gives information about other data and includes descriptive metadata, structural metadata, administrative metadata, reference metadata and statistical metadata.” It is likely that the concepts mapping to the Metadata domain contain data regarding the other procedure_occurrence concepts. Whether they represent descriptive, structural, administrative or reference metadata is unknown. There is also an “NA” standard vocabulary, which only had a concept mapped to the NA domain. R programming language uses NA to represent unavailable or missing values. A possible reason for the presence of NA as both a standard vocabulary and as a concept domain is the presence of concepts which do not map to a standard vocabulary, also defined by OHDSI as non-standard, or source, concepts. In the CDM, only standard concepts are assigned a concept_ID, and the process of converting a source concept to a standard concept is known as mapping. There is a possibility that the procedure_occurrence table included unmapped concepts which were not yet validated by the OHDSI community at the time of data extraction, or if they were already validated by the OHDSI community it is possible that the OMOP teams at each UC health system did not map the concepts. The Book of OHDSI defines a standard concept as “A concept that is designated as valid concept and allowed to appear in the CDM”.

It was surprising to see concepts mapped to the “ICD9Proc” vocabulary. This implies that historical ICD-9 procedure concepts in Epic at one or more of the UC Health systems were not fully mapped to ICD-10 PCS at the time of data extraction. Another observation is that the HCPCS concept domains appear to include “Device”, “Drug”, “Measurement”, “Observation”, and “Procedure”.

It is even more interesting to dive deeper into the concept classes for each domain within the various unique standard vocabularies. Figure 4 demonstrates the varied concept classes and concept domains of concepts within each standard vocabulary in the procedure_occurrence table of UC CORDS. It appears that the standard vocabularies CPT4, HCPCS, ICD10PCS and SNOMED each contain concepts from multiple domains and classes. On the other hand, the Metadata domain within the “None” standard vocabulary only contains concepts with the concept class “Undefined”. This could possibly mean that these concepts are what the Book of OHDSI defines as “classification concepts”.

On re-running my code May 25, 2023 it appears that the data mapping and concept distribution in the SQL query output has now changed. The UC CORDS procedure_occurrence table now consists of only CPT4, HCPCS, ICD-10 PCS, SNOMED and CDT standard vocabularies with just three concept domains which are Procedure, Visit and Provider. The standard vocabularies HCPCS, ICD-10 PCS, SNOMED and CDT only consist of concepts with a Procedure domain. The CPT4 standard vocabulary consists of concepts with Procedure, Visit and Provider domains. There is no longer ICD-9 Procedure standard vocabulary concepts, and there are no longer any concepts mapped to null or “None” standard vocabularies. This demonstrates significant improvement in data quality and data mapping.

When I extracted the UC CORDS data in February 2021, the visit_occurrence table only contained three visit_types and the visit_subtype field was entirely masked. At that time, visit_subtype was still a field not validated by the OHDSI community and therefore it was contained in an extension table that was excluded from the UC CORDS LDS.[46] According to UC CORDS internal documentation for UC Davis users, the reason for including these extension tables in the identified OMOP of UCHDW despite being non-standardized and unvalidated is quoted word-for-word below:

- The UC Health initiative and associated UC Health Data Warehouse has been a key partner and driver of the UCD OMOP instance. As described in other documents, UC Health created OMOP to satisfy both research and operational (administrative) projects and associated use cases
- Many of the use cases for the UC Health Data Warehouse (UCHDW) have been detailed operational projects involving quality and safety and supply chain analytics. These projects required concepts that were not adequately standardized in the OHDSI OMOP model

- To incorporate concepts that are not well captured in the OMOP CDM, UC Health leadership created "Extension" tables to extend beyond the CDM standard[46]

According to the Tableau dashboards available to UCHDW and UC CORDS users, every possible video visit concept from each UC Health site was mapped to the visit subtype “Telemedicine” with standardized code 646370968 as of December 2020, as evidenced by the Table 9, which presents the output from filtering the Code Mapping Explorer for Standardized Code Vocabulary of “Visit Subtype” and “Visit Type”, and Standardized Code Description for “Telemedicine”.[49] A similar search for a Standardized Code Description for “Telehealth” yielded zero results. As you can see in Table 9, there were multiple telemedicine and telehealth source codes at each UC Health site.

Table 9: Code Mapping Explorer output generated from filtering for Standardized Code Vocabulary = “Visit Subtype” or “Visit Type”, and Standardized Code Description = “Telemedicine”.(Paciotti 2020)

Domain	Source Code	Source Code Description	Standardized Code Vocabulary	Standardized Code	Standardized Code Description
UCD Epic	47	TELEMEDICINE	Visit Subtype	646370968	Telemedicine
	76	Telemedicine	Visit Subtype	646370968	Telemedicine
	201	MYC Video Visit	Visit Subtype	646370968	Telemedicine
	2107601	Telemedicine Scheduled	Visit Subtype	646370968	Telemedicine
	2107602	Telemedicine Unscheduled	Visit Subtype	646370968	Telemedicine
	100131001	TELEMEDICINE ACC	Visit Subtype	646370968	Telemedicine
	100131002	TELEMEDICINE J STREET	Visit Subtype	646370968	Telemedicine
	100131003	TELEMEDICINE MAIN HOSP	Visit Subtype	646370968	Telemedicine
	100131004	TELEMEDICINE SHERMAN	Visit Subtype	646370968	Telemedicine
	100131005	TELEMEDICINE CHT	Visit Subtype	646370968	Telemedicine
	100161008	Telehealth	Visit Subtype	646370968	Telemedicine
	100183001	PSYCH TELEMED RANCHO	Visit Subtype	646370968	Telemedicine
UCLA Epic	76	Telemedicine	Visit Subtype	646370968	Telemedicine
	21076	Tele-medicine	Visit Subtype	646370968	Telemedicine
	60812	UCLA Health Telehealth	Visit Subtype	646370968	Telemedicine
	10501207	EMC TELEMEDICINE	Visit Subtype	646370968	Telemedicine
UCSD Epic	40	Telemedicine - Encounter	Visit Subtype	646370968	Telemedicine
	76	Telemedicine	Visit Subtype	646370968	Telemedicine
	319	Telemedicine	Visit Subtype	646370968	Telemedicine
	340	Video Visit	Visit Subtype	646370968	Telemedicine
	7610	Telemedicine (Non-Provider)	Visit Subtype	646370968	Telemedicine
	130190	PMC TELEPAIN	Visit Subtype	646370968	Telemedicine
	190177	MON TELEPSYCH	Visit Subtype	646370968	Telemedicine
	190178	MON TELENEURO	Visit Subtype	646370968	Telemedicine
	200187	MOS TELEHEP	Visit Subtype	646370968	Telemedicine
	200194	MOS TELEENDO	Visit Subtype	646370968	Telemedicine
	220193	MUC TELEONCOLOGY	Visit Subtype	646370968	Telemedicine

	240665	UC HYPERACUTE TELEMED	Visit Subtype	646370968	Telemedicine
	700638	UCSD NICU TELEMEDICINE	Visit Subtype	646370968	Telemedicine
	700639	UCSD PULMONARY TELEMED	Visit Subtype	646370968	Telemedicine
	700640	UCSD GI TELEMED	Visit Subtype	646370968	Telemedicine
	700642	UC Hyperacute Telemedicine	Visit Subtype	646370968	Telemedicine
	700665	HC TELEMEDICINE	Visit Subtype	646370968	Telemedicine
	710230	LJ TELEMEDICINE	Visit Subtype	646370968	Telemedicine
	2406651	UC EMERGENCY TELEMED	Visit Subtype	646370968	Telemedicine
	8003484	Ext Teleradiology	Visit Subtype	646370968	Telemedicine
	8003527	Ext California Protons TeleRadiology	Visit Subtype	646370968	Telemedicine
	8023038	EXT ODC-TELEMED&VSC	Visit Subtype	646370968	Telemedicine
UCSF Epic	76	Telemedicine	Visit Subtype	646370968	Telemedicine
	122	Off License Non-UCSF Telehealth	Visit Subtype	646370968	Telemedicine
	221	Video Visit	Visit Subtype	646370968	Telemedicine
	223	Video Visit Non-Billable	Visit Subtype	646370968	Telemedicine
	5201359	RESPIRATORY CARE CLINIC VIDEO VISIT	Visit Subtype	646370968	Telemedicine
	5201372	UCSF RESPIRATORY CARE CLINIC VIDEO VISIT	Visit Subtype	646370968	Telemedicine
	5201959	PK RESPIRATORY CARE CLINIC VIDEO VISIT	Visit Subtype	646370968	Telemedicine
	5203025	PED HOSPITALIST TELEHEALTH-BIL	Visit Subtype	646370968	Telemedicine
	9020298	NEONATOLOGY TELEHEALTH-BIL	Visit Subtype	646370968	Telemedicine
	9020299	PED CRIT CARE TELEHEALTH-BIL	Visit Subtype	646370968	Telemedicine

As of September 16, 2022 the UC CORDS included a visit_type of "Telehealth".[52] Unfortunately this was well after the date of data extraction for my study. Given that a telehealth visit type is now available, there is scope for future studies to replicate my methodology with the visit_occurrence table and validate both visit_occurrence and procedure_occurrence related concept representations against the telehealth visit type.

DISCUSSION

My study results reflect the lack of standardization in standard vocabulary naming conventions and concept mapping for telehealth. This makes it difficult for researchers to find telehealth-specific data from CDM datasets like UC CORDS, which only capture data mapped to standard vocabularies. My journey through this master's thesis also highlights the multiple data access, data fluency, and data management challenges that clinical researchers face with complex healthcare datasets such as UC CORDS.

Accessing raw, real-world patient data is never a simple endeavor given the privacy, security and ethics issues which can arise. And even after layers of approvals, only skilled analysts can retrieve patient data from the complex data architecture of an EHR like Epic and transform it into an analytic dataset which researchers can utilize. When there is a limited number of analysts to support hundreds of clinical researchers, each with uniquely different data needs, the timeline for obtaining the actual data after approval can take as long as a couple years. This staggered timeline discourages many researchers from carrying out and completing studies which require identifiable data.

The original intent behind the development of UC CORDS was to create an LDS harnessing the power of harmonized data from six health care systems, which would only require a data use agreement and no IRB approval, in turn removing a data access barrier for UC staff and academia to facilitate and encourage more research pertaining to COVID-19. However, it is clear from the diverse mapping in UC CORDS procedure occurrence tables in my results that there are also challenges which come with coordinating data at scale for computable patient phenotypes across multiple sites.

For example, since data in UC CORDS has already gone through two levels of aggregation, researchers accessing UC CORDS may not have much information about the providence of its categorization. It takes a learning curve to understand the structure of an OMOP CDM, and even steeper of a learning curve to understand the tools i.e. SQL, Python, R and platforms i.e. Jupyter required to extract, clean and analyze such data. Additionally, UC CORDS only includes data harmonized to standard concepts. Direct access to source data is highly restricted due to HIPAA concerns. This makes it difficult to visualize data which is not harmonized, for example in this study data on visit subtypes such as "Telehealth" was available in the identified OMOP for hospital operational purposes, but not included in UC CORDS since visit subtypes were not yet harmonized by the OHDSI community.

In light of lessons learned during the pandemic across several sectors, the National Academy of Medicine published “Emerging Stronger After COVID-19: Priorities for Health System Transformation”, with a chapter dedicated to each impacted sector.[53] Each and every chapter of this emphasizes the importance of building a more robust healthcare data infrastructure, in particular the chapters entitled “Digital Health COVID-19 Impact Assessment: Lessons Learned and Compelling Needs” and “Biomedical Research COVID-19 Impact Assessment: Lessons Learned and Compelling Needs”.[54–62] The findings in my thesis support this piece and highlight key recommendations for improving digital infrastructure in preparation for future pandemics.

A study from Columbia Health highlights the importance of involving content experts early when building large data repositories at scale across multiple sites, especially when building a structured data set with computable patient phenotypes.[63] The authors highlight the various stake holders who should be involved in the initial planning and development decisions.[63] They also define desiderata for designing and developing a sustainable model for clinical data infrastructure that allows for patient care as well as clinical research in the most optimum fashion.[63] This is key because while patient care is absolutely vital, it is also important to facilitate clinical research advances and operational analytics for new services by providing a robust digital infrastructure which is tailored to easily cater to each of these functions.

I bring this back to one of my main drivers for performing this study. Clinical researchers who were interested in looking at telemedicine utilization at UC Davis Health, like myself, struggled to extract data on this “new” service due to data harmonization and standardization limitations in the UC CORDS data set and OHDSI community. Most of these telehealth researchers could only access aggregated data from Epic dashboards or spend months to years obtaining ETL outputs from our Data Provisioning Core. If more clinicians interested in telemedicine were involved in OHDSI decisions early on, then telemedicine visit type would have been validated by the OHDSI community earlier and this data would not have been in extension tables of the UCHDW, it would have been incorporated into the standard tables and included in UC CORDS much earlier. This could have opened the door to more telehealth research.

In conclusion, although telemedicine has been considered beneficial for several years, the COVID-19 pandemic offered the best opportunity to improve telemedicine services and fully integrate them into healthcare reimbursement workflows and healthcare information systems. Based on the outcomes of this study, there is still room for process improvement in regard to handling the needs of data capture for new services in emergency scenarios, and healthcare institutions should involve multiple key stakeholders at an earlier stage when developing and implementing a digital infrastructure.

ACKNOWLEDGEMENTS

Graduate Research Mentors and Professional Mentors:

Dr. James Marcin (Chair) and Dr. Prabhu Shankar (Committee), for taking me under their wings as both a graduate student researcher and as a thesis mentee.

Dr. Scott MacDonald, Melissa Jost, Jessie Eting, and Hannah Lytle for welcoming me onto the PEP team as both an extern and a full-time employee.

Each of you have actively shaped my academic, professional, and personal growth throughout the past four years.

Health Informatics Faculty:

Dr. Nick Anderson, Dr. Prabhu Shankar, Prof Mark Carroll (Committee), Dr. Brad Pollock, Dr. Matt Bishop, and Dr. Sean Peisert: for (1) teaching me foundational knowledge and technical skills needed to carry out this master's thesis, and (2) for being an unconditional source of support as mentors and advisors when I needed it the most.

Master of Science in Health Informatics Program Coordinators:

Kelly Byrns and Kristyn Keylon

Pediatric Telemedicine Research Group:

Dr. James Marcin, Dr. Monica Keiko Lieng, Dr. Sarah Haynes for all their support, guidance and mentorship.

University of California Coronavirus 2019 Research Data Set (UC CORDS) Team:

Dr. Brian Paciotti, Steve Covington, Paing Soe for all their help with this data set. I couldn't have carried this study out without their support. Thank you for (1) holding regular weekly office hours and always being available by phone, text, and email to answer any questions I had, and (2) for all your help with data fluency and with constructing my SQL queries! Christy Navarro, librarian informaticist for all her help and support as both an alumni of the program and as a data fluency mentor.

A special thanks to the Center for Data-driven Insights and Innovation at UC Health (CDI2), for its analytical and technical support related to use of the UC Health Data Warehouse and related data assets, including the UC COVID Research Data Set (CORDS).

Family:
My sister, Shayna Kothari, for making me focus/stay on track via Zoom

Classmates, Co-workers, and UC Davis / UC Davis Health Colleagues
Everyone I've met and interacted with has positively impacted my journey. Special thanks to Hawa Kaba and Elsie Bluett, my coworkers and classmates who were an incredible source of support as well.

REFERENCES

1. Lange SJ, Ritchey MD, Goodman AB, et al. Potential Indirect Effects of the COVID-19 Pandemic on Use of Emergency Departments for Acute Life-Threatening Conditions - United States, January-May 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(25):795-800. doi:10.15585/mmwr.mm6925e2

2. Hartnett KP, Kite-Powell A, DeVies J, et al. Impact of the COVID-19 Pandemic on Emergency Department Visits - United States, January 1, 2019-May 30, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(23):699-704. doi:10.15585/mmwr.mm6923e1

3. Kim DKD, Kreps GL. An Analysis of Government Communication in the United States During the COVID-19 Pandemic: Recommendations for Effective Government Health Risk Communication. *World Med Health Policy*. 2020;12(4):398-412. doi:10.1002/wmh3.363

4. Levine CA, Jansson DR. Concepts and Terms for Addressing Disparities in Public Health Emergencies: Accounting for the COVID-19 Pandemic and the Social Determinants of Health in the United States. *Disaster Med Public Health Prep*. June 8, 2021:1-7. doi:10.1017/dmp.2021.181

5. Patel A, Jernigan DB, 2019-nCoV CDC Response Team. Initial Public Health Response and Interim Clinical Guidance for the 2019 Novel Coronavirus Outbreak - United States, December 31, 2019-February 4, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(5):140-146. doi:10.15585/mmwr.mm6905e1

6. 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):1902-1905. doi:10.15585/mmwr.mm6950a4

7. Gostin LO, Friedman EA, Wetter SA. Responding to Covid-19: How to Navigate a Public Health Emergency Legally and Ethically. *Hastings Cent Rep*. 2020;50(2):8-12. doi:10.1002/hast.1090

8. Aquino ER da S, Domingues RB, Mantese CE, Fantini FGMM, Nitrini R, Prado GF do. Telemedicine use among neurologists before and during COVID-19 pandemic. *Arq Neuropsiquiatr*. 2021;79(7):658-664. doi:10.1590/0004-282X-ANP-2020-0488

9. Williams AM, Kalra G, Commiskey PW, et al. Ophthalmology practice during the coronavirus disease 2019 pandemic: the university of pittsburgh experience in promoting clinic safety and embracing video visits. *Ophthalmol Ther*. 2020;9(3):1-9. doi:10.1007/s40123-020-00255-9

10. Zoran S, Turcott C, Whitehead A, Hrabik L, Harris A, Scott Schwoerer J. Rapid Transition to Telemedicine During the COVID-19 Pandemic: Medical Genetics Experience. *WMJ*. 2021;120(3):218-221.

11. Contreras CM, Metzger GA, Beane JD, Dedhia PH, Ejaz A, Pawlik TM. Telemedicine: Patient-Provider Clinical Engagement During the COVID-19 Pandemic and Beyond. *J Gastrointest Surg*. 2020;24(7):1692-1697. doi:10.1007/s11605-020-04623-5

12. Ohannessian R, Duong TA, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. *JMIR Public Health Surveill*. 2020;6(2):e18810. doi:10.2196/18810

13. Gmunder KN, Ruiz JW, Franceschi D, Suarez MM. Factors to Effective Telemedicine Visits During the COVID-19 Pandemic: Cohort Study. *JMIR Med Inform*. 2021;9(8):e27977. doi:10.2196/27977

14. Eichberg DG, Basil GW, Di L, et al. Telemedicine in Neurosurgery: Lessons Learned from a Systematic Review of the Literature for the COVID-19 Era and Beyond. *Neurosurgery*. 2020;88(1):E1-E12. doi:10.1093/neuros/nyaa306

15. Ben-Ari E, Kirschenbaum J, Patel RG, et al. The future of health care service in orthopedic practice: telemedicine or in-person visits? *J Shoulder Elbow Surg*. 2021;30(12):e703-e712. doi:10.1016/j.jse.2021.05.008

16. Pooni R, Pageler NM, Sandborg C, Lee T. Pediatric subspecialty telemedicine use from the patient and provider perspective. *Pediatr Res*. 2022;91(1):241-246. doi:10.1038/s41390-021-01443-4

17. Gross GN. Coding telemedicine visits for proper reimbursement. *Curr Allergy Asthma Rep*. 2020;20(11):73. doi:10.1007/s11882-020-00970-0

18. Hong Y-R, Lawrence J, Williams D, Mainous III A. Population-Level Interest and Telehealth Capacity of US Hospitals in Response to COVID-19: Cross-Sectional Analysis of Google Search and National Hospital Survey Data. *JMIR Public Health Surveill*. 2020;6(2):e18961. doi:10.2196/18961

19. Zhang X, Huang Y, Lee J, Ganta R, Chandawarkar A, Linwood SL. Measuring telehealth visit length and schedule adherence using videoconferencing data. *Telemed J E Health*. 2022;28(7):976-984. doi:10.1089/tmj.2021.0382

20. Jun I, Rich SN, Chen Z, Bian J, Prosperi M. Challenges in replicating secondary analysis of electronic health records data with multiple computable phenotypes: A case study on methicillin-resistant Staphylococcus aureus bacteremia infections. *Int J Med Inform*. 2021;153:104531. doi:10.1016/j.ijmedinf.2021.104531

21. Leung C, Wadhwa H, Sklar M, et al. Telehealth Adoption Across Neurosurgical Subspecialties at a Single Academic Institution During the COVID-19 Pandemic. *World Neurosurg*. 2021;150:e539-e549. doi:10.1016/j.wneu.2021.03.062

22. Maaß L, Pan C-C, Freye M. Mapping Digital Public Health Interventions Among Existing Digital Technologies and Internet-Based Interventions to Maintain and Improve Population Health in Practice: Protocol for a Scoping Review. *JMIR Res Protoc*. 2022;11(3):e33404. doi:10.2196/33404

23. Yu D, He Z. Digital twin-driven intelligence disaster prevention and mitigation for infrastructure: advances, challenges, and opportunities. *Nat Hazards (Dordr)*. 2022;112(1):1-36. doi:10.1007/s11069-021-05190-x

24. Venkatesh A, Malicki C, Hawk K, D'Onofrio G, Kinsman J, Taylor A. Assessing the readiness of digital data infrastructure for opioid use disorder research. *Addict Sci Clin Pract*. 2020;15(1):24. doi:10.1186/s13722-020-00198-3

25. Manny L, Duygan M, Fischer M, Rieckermann J. Barriers to the digital transformation of infrastructure sectors. *Policy Sci*. 2021;54(4):943-983. doi:10.1007/s11077-021-09438-y

26. Van Dort BA, Penm J, Ritchie A, Baysari MT. The impact of digital interventions on antimicrobial stewardship in hospitals: a qualitative synthesis of systematic reviews. *J Antimicrob Chemother*. 2022;77(7):1828-1837. doi:10.1093/jac/dkac112

27. Freeman T, Fisher M, Foley K, et al. Barriers to digital health services among people living in areas of socioeconomic disadvantage: Research from hospital diabetes and antenatal clinics. *Health Promot J Austr*. 2022;33(3):751-757. doi:10.1002/hpja.540

28. Sasaki N, Yamaguchi N, Okumura A, et al. Factors affecting the use of clinical practice guidelines by hospital physicians: the interplay of IT infrastructure and physician attitudes. *Implement Sci*. 2020;15(1):101. doi:10.1186/s13012-020-01056-1

29. Nicolla J, Bosworth HB, Docherty SL, et al. The need for a Serious Illness Digital Ecosystem (SIDE) to improve outcomes for patients receiving palliative and hospice care. *Am J Manag Care*. 2020;26(4 Spec No.):SP124-SP126. doi:10.37765/ajmc.2020.42960

30. Rehm HL, Page AJH, Smith L, et al. GA4GH: International policies and standards for data sharing across genomic research and healthcare. *Cell Genomics*. 2021;1(2). doi:10.1016/j.xgen.2021.100029

31. Alahmar A, Crupi ME, Benlamri R. Ontological framework for standardizing and digitizing clinical pathways in healthcare information systems. *Comput Methods Programs Biomed*. 2020;196:105559. doi:10.1016/j.cmpb.2020.105559

32. Cruz TM, Paine EA. Capturing patients, missing inequities: Data standardization on sexual orientation and gender identity across unequal clinical contexts. *Soc Sci Med*. 2021;285:114295. doi:10.1016/j.socscimed.2021.114295

33. Chapman M, Domínguez J, Fairweather E, Delaney BC, Curcin V. Using Computable Phenotypes in Point-of-Care Clinical Trial Recruitment. *Stud Health Technol Inform*. 2021;281:560-564. doi:10.3233/SHTI210233

34. Onnela J-P. Opportunities and challenges in the collection and analysis of digital phenotyping data. *Neuropsychopharmacology*. 2021;46(1):45-54. doi:10.1038/s41386-020-0771-3

35. Uciteli A, Beger C, Kirsten T, Meineke FA, Herre H. Ontological representation, classification and data-driven computing of phenotypes. *J Biomed Semantics*. 2020;11(1):15. doi:10.1186/s13326-020-00230-0

36. Nour B, Mastorakis S, Mtibaa A. Compute-Less Networking: Perspectives, Challenges, and Opportunities. *IEEE Netw*. 2020;34(6):259-265. doi:10.1109/mnet.011.2000180

37. Koppad S, B A, Gkoutos GV, Acharjee A. Cloud Computing Enabled Big Multi-Omics Data Analytics. *Bioinform Biol Insights*. 2021;15:11779322211035920. doi:10.1177/11779322211035921

38. Brotman JJ, Kotloff RM. Providing Outpatient Telehealth Services in the United States: Before and During Coronavirus Disease 2019. *Chest*. 2021;159(4):1548-1558. doi:10.1016/j.chest.2020.11.020

39. Abbasi-Feinberg F. Telemedicine Coding and Reimbursement - Current and Future Trends. *Sleep Med Clin*. 2020;15(3):417-429. doi:10.1016/j.jsmc.2020.06.002

40. Bajowala SS, Milosch J, Bansal C. Telemedicine pays: billing and coding update. *Curr Allergy Asthma Rep*. 2020;20(10):60. doi:10.1007/s11882-020-00956-y

41. Gausvik C, Jabbarpour Y. COVID-19 Timeline: Centers for Medicare and Medicaid Services (CMS) Changes and Primary Care Support Were Not Enough to Prevent Practice Losses. *J Am Board Fam Med*. 2021;34(Suppl):S7-S9. doi:10.3122/jabfm.2021.S1.200305

42. Yu Y, Zong N, Wen A, et al. Developing an ETL tool for converting the PCORnet CDM into the OMOP CDM to facilitate the COVID-19 data integration. *J Biomed Inform*. 2022;127:104002. doi:10.1016/j.jbi.2022.104002

43. Reese JT, Blau H, Casiraghi E, et al. Generalisable long COVID subtypes: findings from the NIH N3C and RECOVER programmes. *EBioMedicine*. 2023;87:104413. doi:10.1016/j.ebiom.2022.104413

44. Pfaff ER, Girvin AT, Bennett TD, et al. Identifying who has long COVID in the USA: a machine learning approach using N3C data. *Lancet Digit Health*. 2022;4(7):e532-e541. doi:10.1016/S2589-7500(22)00048-6

45. Pfaff ER, Girvin AT, Gabriel DL, et al. Synergies between centralized and federated approaches to data quality: a report from the national COVID cohort collaborative. *J Am Med Inform Assoc*. 2022;29(4):609-618. doi:10.1093/jamia/ocab217

46. Paciotti BM, Soe P. Commonly Asked Questions. covid-lds. September 11, 2020. Accessed February 12, 2021. https://gitlab.ri.ucdavis.edu/ucd-public/covid-lds/-/tree/dev/Commonly%20Asked%20Questions

47. University of California Health (UC Health). University of California COVID Research Data Set (UC CORDS). University of California Health | UC CORDS. Accessed March 30, 2023. https://analytics.uchealth.edu/uccords/

48. University of California Health. University of California Health creates centralized data set to accelerate COVID-19 research. University of California. July 15, 2020. Accessed August 4, 2020. https://www.universityofcalifornia.edu/press-room/university-california-health-creates-centralized-data-set-accelerate-covid-19-research

49. Paciotti BM. Code Mapping Explorer. UCHDW Data Documentation. December 2020. Accessed February 17, 2021. https://tableau.ucdmc.ucdavis.edu/#/site/UCDResearch/views/UCHDWDataDocumentation/CodeMappingExplorer?:iid=1

50. Philofsky M, Olakpe U. Implementing & Adopting a Customized OMOP Common Data Model. *OHDSI*. Published online September 18, 2021. Accessed July 20, 2022. http://www.ohdsi.org/wp-content/uploads/2021/09/18-Implementing-Adopting-a-Customized-OMOP-Common-Data-Model.pdf

51. UC Davis Health. Clinical Data Sources. Health Data Resources. Accessed March 21, 2023. https://health.ucdavis.edu/data/sources.html

52. Paciotti BM, Soe P. UC CORDS: Update Log. covid-lds. September 16, 2022. Accessed September 21, 2022. https://gitlab.ri.ucdavis.edu/ucd-public/covid-lds/-/tree/dev/ATTENTION--Weekly%20Updates

53. National Academy of Medicine; The Learning Health System Series. *Emerging Stronger from COVID-19: Priorities for Health System Transformation*. (Williams A, Lee J, Kadakia K, et al., eds.). National Academies Press (US); 2023. doi:10.17226/26657

54. Lee P, Abernethy A, Shaywitz D, et al. Digital Health COVID-19 Impact Assessment: Lessons Learned and Compelling Needs. *NAM Perspectives*. 2022;2022. doi:10.31478/202201c

55. Cook NL, Lauer MS. Biomedical Research COVID-19 Impact Assessment: Lessons Learned and Compelling Needs. *NAM Perspectives*. 2021;2021. doi:10.31478/202107e

56. DeSalvo K, Hughes B, Bassett M, et al. Public Health COVID-19 Impact Assessment: Lessons Learned and Compelling Needs. *NAM Perspectives*. 2021;2021. doi:10.31478/202104c

57. McClellan M, Rajkumar R, Couch M, et al. Health Care Payers COVID-19 Impact Assessment: Lessons Learned and Compelling Needs. *NAM Perspectives*. 2021;2021. doi:10.31478/202105a

58. Clancy C, Goodrich K, Moody-Williams J, et al. Quality, Safety, and Standards Organizations COVID-19 Impact Assessment: Lessons Learned and Compelling Needs. *NAM Perspectives*. 2021;2021. doi:10.31478/202107d

59. Balser J, Ryu J, Hood M, Kaplan G, Perlin J, Siegel B. Care Systems COVID-19 Impact Assessment: Lessons Learned and Compelling Needs. *NAM Perspectives*. 2021;2021. doi:10.31478/202104d

60. Madara J, Miyamoto S, Farley JE, et al. Clinicians and Professional Societies COVID-19 Impact Assessment: Lessons Learned and Compelling Needs. *NAM Perspectives*. 2021;2021. doi:10.31478/202105b

61. Isasi F, Naylor MD, Skorton D, Grabowski DC, Hernández S, Rice VM. Patients, Families, and Communities COVID-19 Impact Assessment: Lessons Learned and Compelling Needs. *NAM Perspectives*. 2021;2021. doi:10.31478/202111c

62. Mammen M, Narasimhan V, Kuntz R, Lewis-Hall F, Poul M, Schechter A. Health Product Manufacturers and Innovators COVID-19 Impact Assessment: Lessons Learned and Compelling Needs. *NAM Perspectives*. 2022;2022. doi:10.31478/202201b

63. Natarajan K, Weng C, Sengupta S. A Model for Multi-Institutional Clinical Data Repository. *Stud Health Technol Inform*. 2023;302:312-316. doi:10.3233/SHTI230125