



Open camera or QR reader and scan code to access this article and other resources online.

A New Frontier in Telehealth Research: A National Telehealth Data Warehouse

Jason C. Goldwater, MA, MPA,¹ Yunxi Zhang, PhD,² Yael Harris, PhD,¹ Chandra Saurabh,³ and Richard L. Summers, MD⁴

¹University of Mississippi, University Park, Mississippi, USA.

²Department of Data Science, School of Population Health, University of Mississippi Medical Center, Jackson, Mississippi, USA.

Departments of ³Medicine and ⁴Emergency Medicine, University of Mississippi Medical Center, Jackson, Mississippi, USA.

Abstract

Importance: Given the rapid increase in telehealth utilization since the onset of the COVID-19 pandemic, it has become essential to examining the vast amount of available data on telehealth encounters to conduct more cogent, robust, and large-scope research studies to examine the utility, cost-impact, and effect on clinical outcomes that telehealth can potentially provide. However, the diversity of data collected by numerous telehealth organizations has made that type of analysis difficult.

Objective: The University of Mississippi Medical Center (UMMC), a Telehealth Center of Excellence designated by the Health Resources and Services Administration, is creating a National Telehealth Data Warehouse.

Design: UMMC will develop the data warehouse in Microsoft Azure and will use a data dictionary that was created by the Center for Telehealth and eHealth Law (CTeL) to support their national cost-benefit study on the use of telehealth during COVID-19.

Impact: The data warehouse will provide unparalleled opportunities to conduct cost-benefit and cost-effectiveness analyses on telehealth, to develop and test quality measures specific to telehealth, and to understand how telehealth and reduce disparities in health care and expand access to care for

everyone. The warehouse is expected to go live in the Summer of 2023.

Keywords: pandemic, telehealth, telemedicine, commercial telemedicine

Introduction

Standardization of data in the field of telehealth becomes increasingly important as the uptake of telehealth modalities rapidly increases. Although the increased use of telehealth for the past decade is partially attributable to the proliferation of new technologies and changing patient preferences, the COVID-19 pandemic has accelerated telehealth's widespread adoption and increased its acceptability among patients and health care practitioners.¹ In addition to increasing access to physician consults, the pandemic has highlighted the array of services that can be safely and effectively delivered through telehealth including, but not limited to, mental and behavioral health, symptom tracking and monitoring, and disease management.

Through telehealth, many individuals have been able to safely receive health services that were not easily or safely accessible due to the coronavirus.² This significant increase in usage (across a variety of payer types) provides an opportunity to evaluate telehealth's effectiveness in delivering clinical care, its impact on expanding access to care, and its overall value (including cost, quality, efficiency, and outcomes) to public and private payers.

However, most telehealth programs and their institution-based platforms utilize their own data fields and terminologies. This can result in overlooked synonymy and semantic conformities between concepts, producing actionable data that is only applicable to the program from which it was generated.³

To solve this problem, the University of Mississippi Medical Center (UMMC) is creating a National Telehealth Data Warehouse that identifies commonalities across different terminologies and provides a roadmap to articulate a common format with similar classifications that will enable the type of data analysis necessary to inform large-scale policy decisions.

This article discusses the initial conceptualization and structure of the data warehouse and its impact on telehealth data analysis.

DEVELOPING THE DATA DICTIONARY

To create the data dictionary, UMMC examined the format developed by the Center for Telehealth and eHealth Law (CTeL),⁴ which undertook this activity in 2020 to collect data to support their national cost-benefit analysis of telehealth during COVID-19 study that was released in 2021. CTeL gathered data from several synchronous visual telehealth programs that provided the following data elements:

- Date/time of the beginning/end of each telehealth encounter
- ZIP code of originating site
- Diagnoses
- Procedures
- Laboratory orders
- Medication orders
- Insurer
- Amount reimbursed to the health care provider.

Through these initial data fields, the research team was able to identify elements within each data file that could be standardized, as well as perform an initial analysis of the types of services provided, how those services were reimbursed, and additional information, such as the type and number of medications prescribed per encounter.

As shown in *Figure 1*, the initial data dictionary from CTeL was created using the R platform, a free software environment for statistical computing and graphics, and followed three basic steps using the information initially provided by the participating programs.

LINKER DATA FRAME

First, CTeL created a “linker” data frame where each variable description was added and was provided a variable “type” or key. The variable type included all available options that could be provided and generated from the data acquired from those participating programs that volunteered data.

This linker served as an intermediary to build the data dictionary. It contained the names of the variables, a description of each variable provided by each program, and a “variable type.”

MAIN DATA DICTIONARY

The main data dictionary was created using a combined and cleaned data set from all the programs and a linker data frame. Here the data dictionary was built out with variable names, their descriptions, and options.

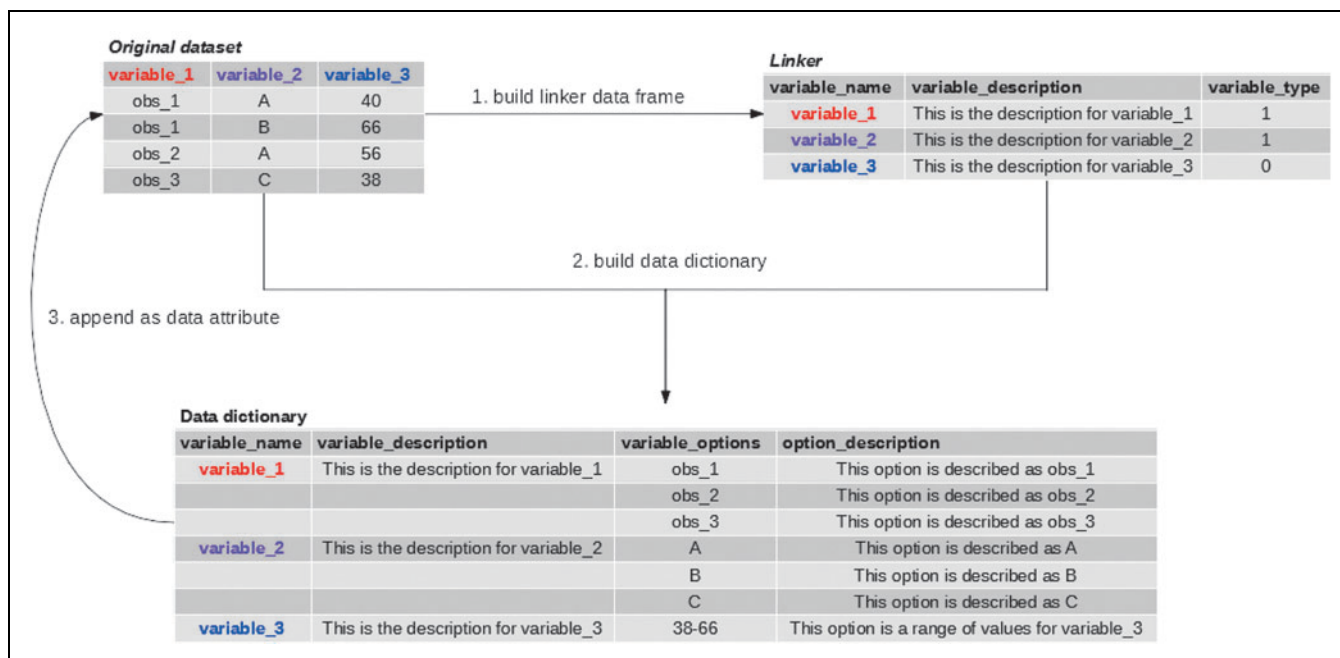


Fig. 1. Developing the Data Dictionary.

Downloaded by University of Mississippi Medical Center from www.liebertpub.com at 07/19/23. For personal use only.

APPENDING AND USING THE DATA DICTIONARY

Finally, the dictionary was appended to the original data sets, along with the date on which the dictionary was created, the author's name, and general attributes included for the data frames to assist with the analysis. For wide distribution, the data dictionary is included as a Microsoft Word table in this document, with an accompanying narrative describing the variables, attributes, and general use of data dictionaries and data frames.

Creating the Data Warehouse

This data warehouse and the analytical system that will support it will bring out the informational potential of telehealth data collected before and during the COVID-19 pandemic as well as afterward. This information provides critical insights to forecast future trends in telehealth, develop quality metrics specific to telehealth, and help evaluate the cost-benefit and cost-effectiveness of telehealth services. Those who participate in contributing data to the warehouse will conform their data files to the standards described in the data dictionary. This will standardize the data across various formats and programs.

DATA WAREHOUSE BUILDING

The development of this data warehouse is a continuous process and represents a complex activity including two major stages. In the first stage, the data warehouse conceptual model is established in accordance with the design. Data sources are then established, as well as the method of extracting and loading data and the storage technology chosen. The National Telehealth Data Warehouse will be created in Microsoft Azure through UMMC and will use external data fields provided by telehealth data programs that conform to the format of the data dictionary. The data will be received by participating in telehealth programs on a regular basis and will be stored with Azure Synapse. This reference architecture implements an extract, transfer, and load (ETL) pipeline that moves data from a UMMC SQL Server database into Azure Synapse.

DEVELOPMENT APPROACH

The data warehouse will follow a *top-down approach* that synergizes the various data files that are submitted and provide a single source of telehealth encounter data for analysis. An iterative approach, which provides a scalable architecture as both the warehouse and user needs grow, is the most adequate method for development.

TESTING AND IMPLEMENTATION

Once the planning and design stages are completed, the testing and implementation stages will commence. UMMC will

develop programs for continually updating and refreshing data, a user interface to allow authorized users access to the data will be developed, sample *ad hoc* queries will be run against the test database, and UMMC will validate the results. In addition, support procedures for data security, back-end recovery, disaster recovery, and data archiving will be implemented.

DEPLOYMENT

The production database is created, and the programs for extracting, validating, transforming, and loading data are run against the source systems to retrieve the telehealth data and store it within the warehouse. UMMC will periodically refresh the warehouse so that users have the most recent information at their disposal. In addition, UMMC will also create a public-facing website that provides information regarding the warehouse in addition to dashboards that provide the latest statistics on telehealth utilization, top diagnoses and procedures used for telehealth, and national demographics on telehealth users.

GOVERNANCE

Access to the warehouse will require potential users to submit a data request with the following: the intent and scope of the research, the specific data elements requested, how the data will be secured, and plans for discarding the data once the research is complete. The request will have to be approved by a governing board led by UMMC, and the request is only valid for the specific scope of the research for which the data have been requested. The warehouse will follow federal guidelines to ensure that data cannot be traced to a specific individual, and access to the data will be free.

Conclusions

Owing to the plethora of virtual visits during the COVID-19 pandemic, the amount of data pertaining to the use of telehealth across clinical areas is significant. This unique project is an extension of the work done by CTeL and will contain a comprehensive standardized data file that has unlimited potential for cogent, detailed, and robust analysis to support and promote telehealth research for the next several decades. It is expected that the data warehouse will launch in the Summer of 2023.

Disclosure Statement

No competing financial interests exist.

Funding Information

This study is supported by the Office for the Advancement of Telehealth, Health Resources and Services Administration,

U.S. Department of Health and Human Services under cooperative agreement award no U6631459.

REFERENCES

1. Roy J, Levy DR, Senathirajah Y. Defining telehealth for research, implementation, and equity. *J Med Internet Res* **2022**;24(4):e35037.
2. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: A systematic review based on current evidence. *BMC Public Health*. **2020**;20(1):1193.
3. Spoladore D, Colombo V, Arlati S, Mahroo A, Trombetta A, Sacco M. An ontology-based framework for a Telehealthcare system to foster healthy nutrition and active lifestyle in older adults. *Electronics* **2021**;10:2129.
4. Goldwater JC, Harris YM. *A cost-benefit analysis of telehealth during COVID-19*. Washington, DC: CTEL, **2021**.

Address correspondence to:
Jason C. Goldwater, MA, MPA
University of Mississippi
University Park 38677
Mississippi
USA

E-mail: jgoldwater@lh-advisors.net

Received: September 29, 2022

Revised: December 16, 2022

Accepted: December 19, 2022

Online Publication Date: February 17, 2023