

In search of longitudinal health data: Bridging the divide between historical medical records and EHRs

Lorrie Dong, MSIS, PhD
School of Information
The University of Texas at Austin

Introduction: Due to current technological and legal limitations, longitudinal studies that require access to historical medical records are extremely difficult, if not impossible, to implement. I am a team member of the *Central State Hospital Digital Library and Archives Project*, a three-year Mellon-funded endeavor to develop technologies and workflows that will improve discovery of and access to digitized historical medical record collections. One goal of the project is to provide the tools to records managers and archivists that will facilitate longitudinal medical, social science, and humanities research based on previously unavailable health sciences data.

Objective: My research goal, and the focus of this poster, is to examine if and how digitized historical medical records can be analyzed in parallel with modern electronic health records (EHRs). If there are comparable categories across these record types, then it would be worthwhile for my project team to develop digital archives features that will allow researchers to easily find and collect historical data that can be used in conjunction with data from EHRs.

Methodology: I examined the contemporary required data fields for certified EHR technology, which are overseen by the Office of the National Coordinator for Health Information Technology, and compared these fields with the categories found in three patient admission registers used by Central State Hospital, a historically African American mental institution in Petersburg, VA, at different points of time from 1868 to 1942.

Results: This table shows the data field requirements for EHR-certified technologies alongside the historical categories from Central State Hospital's admission registers.

Requirement type	Data field	Vocabulary standard	Comparison to Register 1 (1868-1887)	Comparison to Register 2 (1871-1933)	Comparison to Register 3 (1933-1942)
Common Meaningful Use Data Set as defined by Final Rule	Patient name		Name	Name	Name
	Sex		Sex	Sex	Sex
	Date of birth		Age	Age when admitted	Age when admitted
	Race	OMB Ethnicity	[no formal category - "mulatto" notations]	Color	Color
	Ethnicity	OMB Ethnicity			
	Preferred language	ISO 639-2 alpha-3			
	Care team members				
	Medications	RxNorm			
	Medication allergies	RxNorm			
	Care plan				
	Problems	SNOMED-CT	Apparent form of disease	Psychosis	Use of alcohol; Use of drugs; Psychosis; Committed as feeble-minded; Epileptic
	Lab tests	LOINC			
	Lab values/results				Wass. [Wassermann test]
	Procedures	SNOMED-CT or HCPCS/CPT			
Smoking status	SNOMED-CT				
Vital signs					
Criterion-specific data requirements for EHR certification	Provider name & contact info				
	Reason for referral				
	Encounter diagnoses	SNOMED CT or ICD-10-CM	Supposed cause of lunacy; Condition when discharged; Cause of death	Assigned cause of psychosis; Condition when discharged; Cause of death	Condition when discharged; Cause of death
	Cognitive status			Psychosis; Blind, deaf, or dumb; Epileptic or paralytic	Psychosis; Blind, deaf, or dumb; Committed as Feeble-minded; Epileptic
	Functional status			Crippled or deformed; Blind, deaf, or dumb; Epileptic or paralytic	Blind, deaf, or dumb; Paralysis, crippled, or deformed
	Discharge instructions				
Immunizations	CVX				

Findings: Of the 23 EHR data fields, only nine corresponded with historical admission categories. Unsurprisingly, the number and types of categories for patient records, as well as the record media and formats, changed over time. I nevertheless found it would be possible to analyze longitudinally certain information. For example, a researcher could examine data related to "Race," an EHR category, and "Color," a legacy category. Similar comparisons could be made between the modern "Encounter diagnoses" and the historical "Assigned cause of psychosis" or "Condition when discharged."

The potential to compare medical data longitudinally comes with a caveat. In both examples that I provided, the categories over time (including across the registers) are *not* parallel to one another in terms of name, definition, and medical or administrative purpose. Therefore, meaningful analysis by researchers in any field will require an understanding of social and historical context.

Future Research: This research demonstrates the value of creating retrieval tools for digital archives systems that allow users to find and examine historical health classification data in relation to EHR data. In addition to the work to develop these tools, it would be useful to study the potential parallels between historical medical record formats and EHR document templates, of which there are currently nine; this may facilitate the development of search and analysis features based on formatting rather than character recognition, thus by-passing some of the challenges of working with hand-written and/or privacy-restricted materials.