

Children's Mercy Kansas City

## SHARE @ Children's Mercy

---

Manuscripts, Articles, Book Chapters and Other Papers

---

12-2019

### Heterogeneity introduced by EHR system implementation in a de-identified data resource from 100 non-affiliated organizations.

Earl F. Glynn

*Children's Mercy Hospital*

Mark A. Hoffman

*Children's Mercy Hospital*

Let us know how access to this publication benefits you

Follow this and additional works at: <https://scholarlyexchange.childrensmercy.org/papers>



Part of the [Health Services Research Commons](#)

---

#### Recommended Citation

Glynn EF, Hoffman MA. Heterogeneity introduced by EHR system implementation in a de-identified data resource from 100 non-affiliated organizations. *JAMIA Open*. 2019;2(4):554-561. Published 2019 Aug 7. doi:10.1093/jamiaopen/ooz035

This Article is brought to you for free and open access by SHARE @ Children's Mercy. It has been accepted for inclusion in Manuscripts, Articles, Book Chapters and Other Papers by an authorized administrator of SHARE @ Children's Mercy. For more information, please contact [hlsteel@cmh.edu](mailto:hlsteel@cmh.edu).

## Research and Applications

# Heterogeneity introduced by EHR system implementation in a de-identified data resource from 100 non-affiliated organizations

Earl F Glynn<sup>1</sup> and Mark A Hoffman<sup>1,2,3</sup>

<sup>1</sup>Children's Mercy Hospital, Children's Research Institute, Kansas City, Missouri, USA and <sup>2</sup>Department of Pediatrics, University of Missouri Kansas City, Kansas City, Missouri, USA and <sup>3</sup>Department of Biomedical and Health Informatics, University of Missouri Kansas City, Kansas City, Missouri, USA

Corresponding Author: Mark Hoffman, PhD, Chief Research Information Officer, Children's Mercy, Kansas City, MO, USA; mhoffman@cmh.edu

Received 14 June 2019; Editorial Decision 28 July 2019; Accepted 1 August 2019

### ABSTRACT

**Objectives:** Electronic health record (EHR) data aggregated from multiple, non-affiliated, sources provide an important resource for biomedical research, including digital phenotyping. Unlike work with EHR data from a single organization, aggregate EHR data introduces a number of analysis challenges.

**Materials and Methods:** We used the Cerner Health Facts data, a de-identified aggregate EHR data resource populated by data from 100 independent health systems, to investigate the impact of EHR implementation factors on the aggregate data. These included use of ancillary modules, data continuity, International Classification of Disease (ICD) version and prompts for clinical documentation.

**Results and Discussion:** Health Facts includes six categories of data from ancillary modules. We found of the 664 facilities in Health Facts, 49 use all six categories while 88 facilities were not using any. We evaluated data contribution over time and found considerable variation at the health system and facility levels. We analyzed the transition from ICD-9 to ICD-10 and found that some organizations completed the shift in 2014 while others remained on ICD-9 in 2017, well after the 2015 deadline. We investigated the utilization of "discharge disposition" to document death and found inconsistent use of this field. We evaluated clinical events used to document travel status implemented in response to Ebola, height and smoking history. Smoking history documentation increased dramatically after Meaningful Use, but dropped in some organizations. These observations highlight the need for any research involving aggregate EHR data to consider implementation factors that contribute to variability in the data before attributing gaps to "missing data."

**Key words:** electronic health record, phenotype, data science, data visualization

## BACKGROUND AND SIGNIFICANCE

Major advances in genomics during the past two decades have generated a growing body of genotypic data. Beyond Mendelian conditions, cancer analysis and some pharmacogenetic associations, the subtle associations between genotype and phenotype are still poorly understood. In order to gain deeper insights into human health and disease, richly annotated phenotypic data is required to understand

genotype-phenotype relationships. Electronic health records (EHR) are recognized as a primary source of data for digital phenotyping.<sup>1–5</sup> EHR systems are used to document clinical observations in nursing and physician notes, record laboratory findings, automate medication ordering, provide a platform for complex decision support and integrate financial and clinical operations. Each of these processes generate machine-readable data that can be consumed by

analytical applications, a process that has been described as “Evidence generating medicine.”<sup>6</sup> EHR systems also serve as the legally binding medical record and must support compliance with a variety of regulatory frameworks including the Health Insurance Portability and Accountability Act (HIPAA). The Meaningful Use initiative of the 2009 American Recovery and Reinvestment Acts provided funding to stimulate the adoption of EHR systems, resulting in the widespread availability of machine-readable health information.<sup>7</sup>

Many EHR-based phenotyping projects have used EHR data from single organizations to demonstrate the potential for digital phenotyping. For example, the Vanderbilt Synthetic Derivative has been used to examine hypertension and other clinical conditions.<sup>8</sup> Beth Israel Deaconess Medical Center releases a subset of their EHR data through the MIMIC initiative.<sup>9</sup> The MIMIC-III data has been used to evaluate discharge summaries for 10 phenotypes in support of a comparison of deep learning and concept extraction methods.<sup>10</sup> These projects are facilitated by the ability of the investigators to negotiate with a single legal entity to gain access to the data. Investigators can contact the team involved in the implementation of the EHR to understand workflow and data decisions. Data from these single-organization projects tends to be harmonized. Key limitations of this approach include limited numbers of patients and a potential lack of geographic and demographic diversity. One solution to these limitations is to integrate EHR data from disparate sources in multiple regions.

A few initiatives, including the electronic medical records and genomics (eMERGE) network and PCORNET, the National Patient-Centered Clinical Research Network, support decentralized analyses across disparate systems by distributing standardized queries to member sites.<sup>11,12</sup> The PCORNET sites have agreed to use a Common Data Model and often process these queries through standardized local implementations of i2b2.<sup>12</sup> Another model involves the extraction of a subset of EHR data from participating sites and loading of the data into an aggregate data warehouse. In some cases, these warehouses are supported by major EHR vendors. This model has been used by GE through their Medical Quality Improvement Consortium and Cerner through their Health Facts (HF) initiatives.<sup>13,14</sup> Epic, another major EHR vendor, has referenced an aggregate data resource, Cosmos, but the system is not yet available in production.

Benefits of working with these large-scale data resources include the high numbers of patients and facilities covered and the diversity of participating organizations and their patient populations. For example, HF has been operational since 2000 and includes data from nearly 69 million patients treated at 664 facilities associated with 100 nonaffiliated health systems throughout the US. Research using HF has been used to support a variety of outcomes research, including evaluations of the impact of potassium and magnesium levels on risks associated with myocardial infarction.<sup>15,16</sup> Other work has investigated factors affecting readmission risks.<sup>17,18</sup> The benefits of this resource are offset by a high level of heterogeneity and gaps in the data. For research focused resources, such as Cerner HF, the data is fully de-identified to HIPAA standards and contributing facilities are masked, preventing researchers from seeking clarifications about missing or aberrant data or implementation decisions. Some researchers working with these data resources may have limited understanding of EHR systems and may assume uniform implementation of the EHR across contributing organizations when performing analyses. Each implementation of the systems from a common EHR vendor may make significantly different use of the

same set of core features and capabilities. Most EHR systems offer modules that support ancillary sections such as the pharmacy, laboratory, emergency department, surgery, and radiology. Some organizations may not implement all ancillary modules or may activate modules in a phased manner. Likewise, the same clinical finding or outcome may be documented in many different ways in the system. These variations can have considerable impact on the calculation of the denominators used in digital phenotyping studies.

We demonstrate methods and strategies to evaluate heterogeneity between contributors to a de-identified, multiorganization, EHR data warehouse in order to promote more accurate analysis.

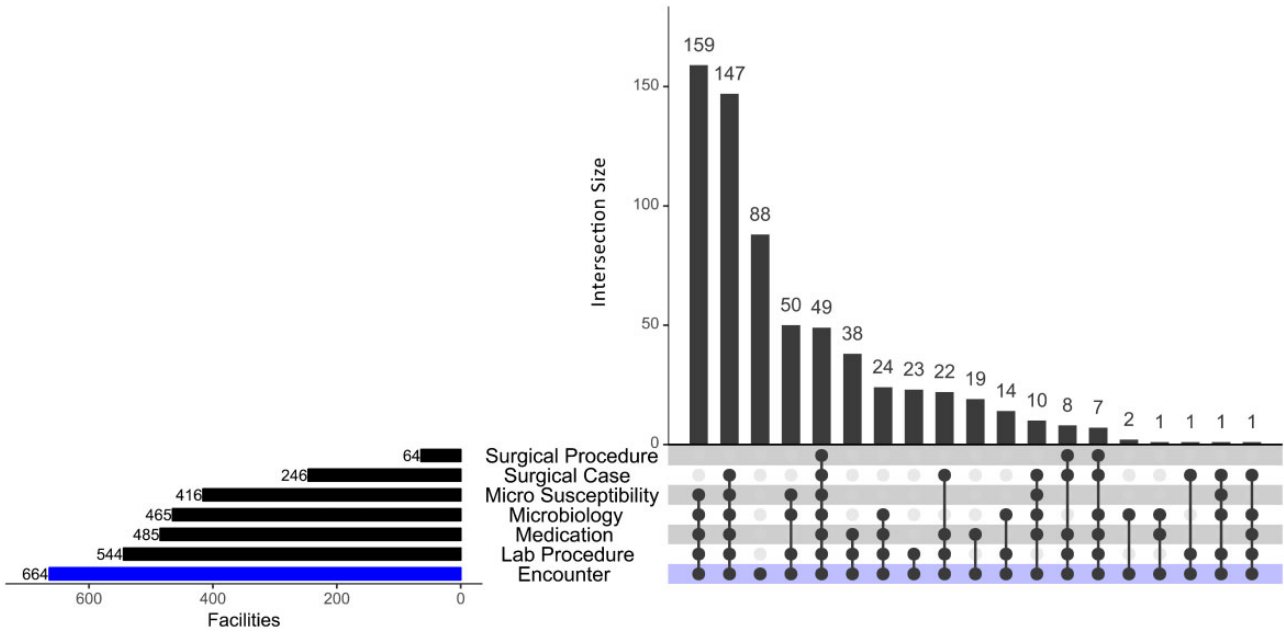
## METHODS

Cerner HF is populated by the daily extraction of discrete EHR data from participating organizations. These organizations have provided data rights to Cerner and allow the integration of de-identified information into the data warehouse. HF data is de-identified to HIPAA standards; text documents and images are not included. Children’s Mercy is a contributor to HF and has received a copy of the full database to support research. The data is installed in Microsoft (Redmond, WA) Azure and queries are performed with Microsoft SQL Server Management Studio version 17.9 and R Studio version 1.1.453 with R version 3.5.1.<sup>19</sup> This work was performed with the 2018 version of the HF with data from 2000 through 2017. Data from 664 facilities associated with 100 nonaffiliated health systems are included in this release. This version of the HF data includes 69 million patients, 507 million encounters, 4.7 billion lab results, 729 million medication orders, 989 million diagnoses, and 6.9 billion clinical events. The Children’s Mercy Institutional Review Board has designated research with HF data as “non-human subjects research.”

Every patient encounter in HF is associated with a facility and health system ID. These IDs are masked and are not linked to the name of the contributing organization. For purposes of this description, a health system is a legal entity including one or more facilities. Results and transactions in HF are stored in fact tables while dimension tables provide descriptive content. Encounter IDs that can be used to join to an encounter tables which in turn can be connected to a dimension table that provides information about the de-identified hospital and health system.

## Module adoption

Many EHR systems are comprised of modules that support specific clinical processes or units such as the pharmacy, laboratory, or surgery. Some organizations may use the core EHR and all available ancillary modules from the same vendor. Other organizations may have a laboratory system provided by a vendor distinct from their core EHR. Ambulatory facilities using an EHR are not likely to need an inpatient pharmacy or laboratory module. HF includes substantial laboratory, inpatient pharmacy, and surgical data. The laboratory data in HF can include clinical pathology or “General Lab” and microbiology, which is divided into the culture results and antibiotic susceptibility results. The surgical data includes cases and procedures. Each of these categories has a separate fact table in HF. Data from other Cerner modules, including anatomic pathology and the emergency department, are not included in the current release of HF.



**Figure 1.** In this UpSet graph, each row represents a module in Cerner. Each column represents a combination of modules. A filled dot indicates that the module is a member of that particular set.

**Continuity of data feeds**

HF has been operational for nearly 20 years. Every day an extract process runs at each participating organization and de-identified data is transferred to the Cerner data center for incorporation into HF. Especially during the early years of operation, a disruption in the data feed may not always have been detected for some time. This causes temporal gaps in the data that vary from facility to facility. In order to identify these gaps, we scanned the encounter volume at each contributing health system and facility as a function of time.

**Shift in coding systems for diagnoses**

The International Classification of Disease (ICD) is widely used to document diagnoses and procedures in support of billing and other administrative processes. The ICD-9 version of this coding system has been in widespread use to document diagnoses for most of the time that HF has been operational. The Center for Medicare and Medicaid Services required the use of ICD-10 by October of 2015. We evaluated the prevalence of ICD-9 versus ICD-10 diagnosis codes across contributing facilities by querying the billing data by encounter.

**Representation of death**

Determining morbidity and mortality from EHR data can be challenging. Death is the most significant outcome to monitor but is surprisingly difficult to ascertain from discrete EHR data. In HF, death is primarily documented in the discharge disposition, an attribute of the encounter record.<sup>20</sup> In order to determine how consistently this field is used to document death, we evaluated the four discharge disposition codes indicative of death, “expired,” “expired at home. Medicaid only, hospice,” “expired in a medical facility. Medicaid only, hospice,” and “expired, place unknown. Medicaid only, hospice.” We grouped encounters by the relative size of the facility, using the range of the number of beds as a proxy for size.

**Representation of discrete clinical findings**

HF uses a general-purpose table, “Clinical Events,” to store a wide variety of clinical findings and other information that are not captured in the topic-specific fact tables. In the version of the data that we used there are 669 distinct clinical events. We evaluated a less common set of prompts related to travel to Africa in response to the Ebola outbreak in 2014.<sup>21</sup> We chose height as a representative common measurement with many variations and queried the clinical event table to evaluate how many health systems and facilities use the “Height-Length” group, part of the “Vital Sign” category. Compliance with the first phase of Meaningful Use funding required documentation of smoking history,<sup>22</sup> also stored in the clinical events table.

**RESULTS**

The module adoption by facility and organization was evaluated using an UpSet diagram (Figure 1), a useful format for evaluating overlapping sets.<sup>23</sup> We found that of the 664 facilities in HF, 49 contributed to all 6 of the module-associated fact tables. In contrast, 88 facilities did not contribute to any of them. The largest group (159 facilities), provided data for all categories except the two associated with surgery. There were 49 facilities that populated the Microbiology table but not the Micro Susceptibility table. In Cerner systems there are two options for receiving a microbiology isolate—through the Cerner PathNet Microbiology module, which populates both the HF Microbiology table and the Micro Susceptibility tables, and through an interface to a non-Cerner laboratory system. The latter only populates the Microbiology table, not the susceptibility table. The Cerner surgery module consists of two key features—the case and the procedures. There are more organizations using the surgical case feature than the surgical procedures, which offer very granular visibility into time stamped events during surgery.

We speculated that facilities with 5 beds or fewer, a useful proxy for ambulatory sites, would be less likely to have implemented the



**Figure 2.** (A) Each spark graph represents encounters in one-month intervals between 2000 and 2017 for a distinct health system among the 100 health systems contributing to Health Facts. The left of each spark graph is the year 2000 and the right represents 2017. The height of each spark graph represents relative encounter activity. The scale of each health system varies. (B) Spark graphs for two representative health systems are expanded to show corresponding facility spark graphs. Each Spark graph represents monthly encounters. The left of each spark graph is the year 2000 and the right represents 2017. The height of each spark graph represents relative encounter activity.

ancillary modules. We found that of the 294 facilities with 5 or fewer beds, only 84 did not use any of the ancillary modules, while 210 had evidence that they use one or more ancillary module.

We evaluated the continuity of the HF data feeds by examining the number of encounters over time. When reviewed at the health system level (Figure 2A), it is clear that some facilities contributed early in the life cycle of HF but stopped. Others began contributing more recently. Some began in the middle of the life cycle and then terminated. In many instances there are brief drops in data contribution and then the feed was restored. One health system (second row, fourth column) has an unusual cyclical pattern. Preliminary analysis of this health system for variable patterns in diagnosis code utilization did not provide a feasible explanation. We reviewed consistency between facilities in a common health system. Two representative health systems are shown in Figure 2B, one in which there was general consistency between five facilities (the left column) and one in which there is variation between eight facilities (the right column).

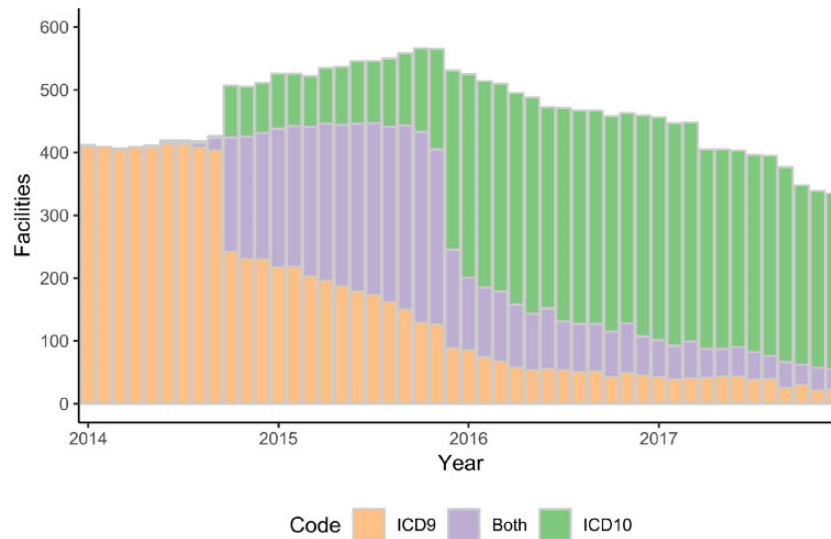
The timing of the shift from ICD-9 to ICD-10 varies. Figure 3 shows that some organizations converted their diagnosis coding to ICD-10 in late 2014 while others had not yet completed the transition in 2017, the latest period available in the version of HF that we used.

We investigated variation in documentation for a fundamental outcome indicator, death, as documented using a discharge disposition with one of four variations of “expired” (Table 1). In order to qualify a facility had to have one or more encounter in which the discharge disposition included “expired.” For this study we did not pursue relative numbers of expired patients. Two facilities in the data were not classified by bed count, one of these used expired in the discharge disposition while the other did not. We found that 23% of smaller ( $\leq 5$  bed) facilities had a history of using the discharge disposition of expired. Medium-sized facilities from 6 to 199 beds were most likely to document a discharge disposition of death. Larger facilities were more likely to use the discharge disposition of

death but for each category there was a considerable percentage of sites that are not documenting deaths in this manner.

Two clinical findings, travel status and height-length, were examined to explore how commonly they were contributed to the data (Table 2). Ten health systems with 45 facilities incorporated travel history related to visiting Africa. Among the documentation options, some of these have differing but overlapping levels of precision, for example “travel to Africa within last 21 days” and “travel to Africa within the last month” are conceptually similar while “Travel to Guinea/Liberia/Sierra Leone” is more specific about the location but less precise about the timing. We found that 73 health systems and 485 facilities documented a variation of “Height-Length” a total of 127 996 345 times. In order to understand the variations in how these organizations document height-length, we examined the 10 distinct event descriptions used in this category as shown in Table 2. The generic description “Height” was used by the most facilities, 477, followed by “Height, Estimated,” used by 247 facilities. Height in inches was observed more often than estimated height, but at fewer facilities. Despite the naming convention for “Height in inches,” the actual unit of measure associated with the findings varied among the organizations (data not shown).

We evaluated the Meaningful Use driven documentation of smoking history. There are 25 unique clinical events in the “Smoke/Tobacco” group of the “Substance Use” category of clinical events in HF. The most widely utilized prompts were “Tobacco type” with 11 596 602 uses and “Smoke, Exposure to Tobacco Smoke,” with 9 358 048 uses. The third most utilized was “smoking history” with 8 756 872 uses. Unlike smoke exposure, which could include second hand smoke, and tobacco type, which could reflect a secondary reflex prompt, “smoking history” seemed to be a patient-specific primary question. There are 18 health systems with 59 facilities using a “smoking history” clinical event. Of these, four systems had fewer than 100 events and were excluded from further analysis. As expected, there was a sharp increase in use of this prompt after



**Figure 3.** Transition from ICD-9 to ICD-10. The height of each bar represents the number of facilities providing diagnosis codes per year, the color indicates which coding system was in use, including some facilities that used both ICD-9 and ICD-10 in a given year.

**Table 1.** Frequency of documenting death using discharge disposition of “expired”

Facility bed size (range)	Number of facilities	Facilities reporting deaths	Facilities not reporting deaths	% Reporting deaths
1–5	294	67	227	22.8%
6–99	154	141	13	91.6%
100–199	80	72	8	90.0%
200–299	63	48	15	76.2%
300–499	43	36	7	83.7%
500+	28	19	9	67.9%
Unknown	2	1	1	50.0%
All facilities	664	384	280	57.8%

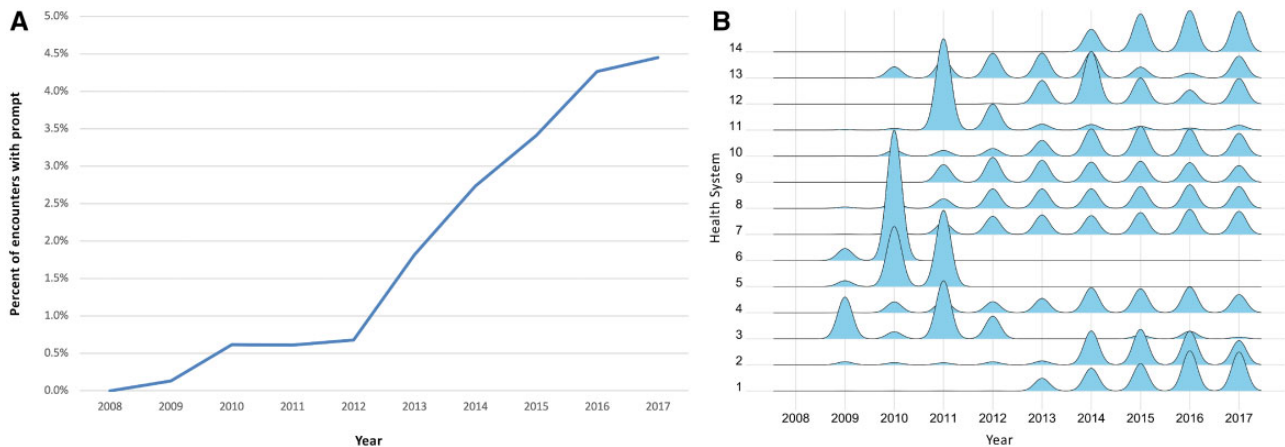
**Table 2.** Health system and facility usage of representative clinical events

Category	Group	Event description	Health systems	Facilities	Unique records
Travel	Travel location	African travel location	1	6	27
Travel	Travel location	Recent travel to West Africa	1	6	3 215 896
Travel	Travel location	Travel to African country	5	24	1 235 487
Travel	Travel location	Travel to West Africa within last month	1	1	92 708
Travel	Travel timing	Travel to Africa within the last 21 days	1	4	1154
Travel	Travel timing	Travel to Guinea/Liberia/Sierra Leone within past year	1	4	722
Vital sign	Height–Length	Height	72	477	103 411 085
Vital sign	Height–Length	Height, Body Surface Area	1	2	221
Vital sign	Height–Length	Height, Estimated	45	247	3 591 945
Vital sign	Height–Length	Height, Feet	12	48	4 756 482
Vital sign	Height–Length	Height, Inches	31	174	14 434 466
Vital sign	Height–Length	Height, Measured	8	17	864 436
Vital sign	Height–Length	Height, Percent	14	41	11 365
Vital sign	Height–Length	Height, Percent for age	16	44	4918
Vital sign	Height–Length	Length, Birth	52	238	907 501
Vital sign	Height–Length	Length, Infant	5	20	13 926

2009(Figure 4A). When we investigated use of “Smoking History” at the individual health system level (Figure 4B). The ridgeline plots are scaled by each system, for example ridgeline 1 represents a total of 4.7 million events while ridgeline 2 represents 194 000. We noted that four (3, 5, 6, and 11) of the 14 systems seem to have used the prompt initially and then discontinued use

in subsequent years. Systems 3 and 11 have a lower but continuing level of usage while Systems 5 and 6 have zero usage from 2012 to 2017. Given the variability in data contribution over time, we verified that all four of these systems continued to have active data feeds during the years with low usage of the smoking history prompt.





**Figure 4.** (A) Use of smoking history prompt. For the 14 health systems with more than 100 encounters that used the smoking history prompt, we express the number of smoking history prompts per year as a percentage of total encounters in the same year. (B) Annual smoking history documentation by health system. Ridges plot showing frequency of usage of smoking history by health system as a function of time. Each ridgeline is a density plot for the events of the indicated health system. Each plot is scaled by the data from that health system.

## DISCUSSION

Aggregate EHR data has tremendous potential to support digital phenotyping, population level analyses, comparisons of provider behavior at disparate organizations, outcomes research, and quality improvement. Investigators working with aggregate EHR data are not always familiar with the nuances of EHR implementations and the significant impact that they can have on the accuracy of a data analysis. In this article, we provide representative examples that can inform the careful design of analyses that use multisite, de-identified EHR data. We used the availability of categories of data to infer implementation decisions at sites contributing data to Cerner HF.

Organizational purchasing decisions, including the selection of optional EHR modules, impacts the availability of key categories of data. We found that the implementation of ancillary modules, including laboratory, pharmacy and surgery, and varies across organizations. We also noted that the use of these modules differs, for example, many organizations use the surgical case feature but not the surgical procedures feature. Based on this finding, one design consideration for data analyses should be whether or not a given category of data necessary for a project is available for each facility in the data set. If the data is not available, those facilities should be excluded from that part of the analysis.

HF has been operational for nearly 20 years. As with many large-scale initiatives, some early participants withdrew or changed EHR vendor, later participants joined and there have been occasional drops in the connection of a data feed. We found substantial variation across time in the data feeds at both the inter-health system and intra-health system levels. The intra-health system variation in data continuity could be due to implementations in which the installation was not centralized, the opening of a new facility, closure of an existing facility or a merger of previously unaffiliated organizations. The timing of data contribution can have significant impact on analyses that are time sensitive, whether public policy analyses or research related to medications that were not available during segments of the time available in the data set.

Our analysis of the shift from ICD-9 to ICD-10 shows that this migration has not been uniform among the 664 facilities covered in this analysis. Indeed, despite the deadline in October of 2015, we

noted several organizations continued to use ICD-9 in 2017 while others appear to have completed their migration in 2014. Several studies have shown a shift in the observed rates for conditions including psychoses<sup>24</sup> and injuries requiring hospitalization<sup>25</sup> before and after the adoption of ICD-10. From the analysis standpoint, this requires users of aggregate data to give careful consideration to which version of ICD is in use at each organization during the time period being evaluated and verifies that assuming 1 October 2015 as a hard date for the shift would result in inaccurate analyses. While several other evaluations of the ICD-9 to ICD-10 shift have been performed for individual organizations<sup>25,26</sup> and in a study of 13 systems,<sup>27</sup> our analysis of 100 nonaffiliated health care systems provides a coverage of the change in a very large and diverse set of health care facilities.

Aggregate EHR data has significant value for outcomes research. Understanding factors that contribute to mortality is a key goal of many outcomes research projects, yet death is not consistently documented in EHR systems. We evaluated the most common field used to document death in the Cerner EHR, discharge disposition. Notably, small to medium facilities are the most likely to use a discharge disposition of “expired” while only 68% of the facilities with 500 or more beds used the field to document death. There are alternative means to document death in an EHR, including diagnosis codes and in textual clinical documentation. Death-related diagnosis codes are found infrequently in HF and text notes are not available because they cannot be reliably de-identified. Confirmation of death can be accomplished by linking to the Social Security master death index but the de-identified nature of aggregate EHR data prevents this.

We demonstrate that discrete documentation can vary among organizations and over time. Travel history is available from a small group of health systems and there are differing levels of precision in how the history is captured. In contrast, heights are available from many facilities but there are wide variations in the specific expression of the data capture field. Some fields became available considerably later than initial participation in HF or the initial EHR implementation. One example, smoking history, was driven by regulatory factors. We saw a strong increase in the use of this prompt following the 2009 Meaningful Use funding. Yet, the overall utilization of the prompt seems to be low (less than 5%). A public health

**Table 3.** Recommendations for aggregate EHR data analysis

Source of variation	Recommendation
Variation in ancillary module adoption	Evaluate facility level use of topic specific data tables and fields.
Temporal variation of contributors	Evaluate contribution of each organization over time. Structure analyses to exclude organizations when they were not actively contributing data.
ICD version	For analyses that span the transition period from ICD-9 to ICD-10, estimate when data contributing organizations shifted from ICD-9 to ICD-10.
Outcome measure	For each organization, evaluate the availability of key outcomes measures. Exclude organizations that are not documenting the measure in the expected manner. Do not assume continuous, consistent usage.
Variation in documentation	Confirm that each organization in an analysis is using the documentation prompt(s) needed for an analysis, exclude those that do not. Evaluate conceptual overlaps in the descriptions of variations for similar concepts. Evaluate whether usage of prompt(s) is consistent over time.

researcher unaware of the regulatory factors affecting documentation may draw mistaken conclusions about smoking during this time period. Interestingly, at the individual health system level, some organizations discontinued use of the prompt. They may have changed to another prompt or ceased to document smoking after a certain period. This observation raises a general issue relevant for both aggregate and local data warehouse analyses—usage of discrete prompts changes over time but researchers and analysts may not always be informed of these changes.

We demonstrate that analysis based on aggregate EHR data must include inter-facility variation as a consideration in the analysis plan. This variation should be addressed based on the specific data element(s) required for the analysis. For example, a project evaluating mortality between 2005 and 2010 related to an adverse drug reaction that could have been avoided based on a laboratory result would require the exclusion of facilities that do not use the pharmacy module, that lack the laboratory module, that do not use discharge disposition of “expired” and/or were inactive between the required dates. The inclusion of any of these facilities with implementation-based data gaps would potentially lead to an incorrect denominator in the analysis. A project using diagnosis codes to evaluate co-morbidities would need to address the shift from ICD-9 to ICD-10. In Table 3, we offer recommendations that will inform future work with aggregate EHR data analyses.

Aggregate EHR data offers a valuable resource to inform many types of research and quality improvement work. Some data issues that have been attributed to “missing data” have their origins in system implementations and are justified based on the local use of the contributing EHR system. Our work clearly demonstrates that aggregate EHR data is not uniform at the temporal or content levels. The assumption that locally implemented systems from a common EHR vendor are standardized is not supported by this work. The data science methods used in this analysis can readily be applied to other EHR data types that are affected by local implementation or purchasing decisions. Our examples are representative of wider

issues with similar manifestations, for example variations in workflow such as reflex ordering of some lab tests, changes in clinical leadership and other factors can have similar effects on the data. In order to work effectively with aggregate EHR data, it is important to understand the attributes and characteristics of EHR implementations and to incorporate methods that account for inter-facility variations in the analysis plan.

## ACKNOWLEDGMENTS

Sierra Martin and Shivani Sivasankar for helpful discussions and comments. Matthew Breikreutz for technical assistance with figures.

## FUNDING

This work was supported by a MCA Partners Advisory Board grant from Children’s Mercy Hospital (CMH) and The University of Kansas Cancer Center (KUCC).

## CONFLICT OF INTEREST STATEMENT

None declared.

## REFERENCES

1. Son JH, Xie G, Yuan C, *et al*. Deep phenotyping on electronic health records facilitates genetic diagnosis by clinical exomes. *Am J Hum Genet* 2018; 103 (1): 58–73.
2. Albers DJ, Elhadad N, Claassen J, *et al*. Estimating summary statistics for electronic health record laboratory data for use in high-throughput phenotyping algorithms. *J Biomed Inform* 2018; 78: 87–101.
3. Chiu PH, Hripcsak G. EHR-based phenotyping: bulk learning and evaluation. *J Biomed Inform* 2017; 70: 35–51.
4. Denny JC, Bastarache L, Roden DM. Phenome-Wide Association Studies as a tool to advance precision medicine. *Annu Rev Genomics Hum Genet* 2016; 17 (1): 353–73.
5. Wei WQ, Denny JC. Extracting research-quality phenotypes from electronic health records to support precision medicine. *Genome Med* 2015; 7 (1): 41. <https://www.ncbi.nlm.nih.gov/pubmed/25937834>.
6. Embi PJ, Payne PR. Evidence generating medicine: redefining the research-practice relationship to complete the evidence cycle. *Med Care* 2013; 51 (8 Suppl 3): S87–91.
7. Adler-Milstein J, Jha AK. HITECH Act drove large gains in hospital electronic health record adoption. *Health Aff (Project Hope)* 2017; 36 (8): 1416–22.
8. Teixeira PL, Wei WQ, Cronin RM, *et al*. Evaluating electronic health record data sources and algorithmic approaches to identify hypertensive individuals. *J Am Med Inform Assoc* 2017; 24 (1): 162–71.
9. Johnson AE, Pollard TJ, Shen L, *et al*. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016; 3: 160035.
10. Gehrmann S, Dernoncourt F, Li Y, *et al*. Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives. *PLoS One* 2018; 13 (2): e0192360.
11. Kho AN, Pacheco JA, Peissig PL, *et al*. Electronic medical records for genetic research: results of the eMERGE consortium. *Sci Transl Med* 2011; 3 (79): 79re1.
12. Waitman LR, Aaronson LS, Nadkarni PM, *et al*. The Greater Plains collaborative: a PCORnet Clinical Research Data Network. *J Am Med Inform Assoc* 2014; 21 (4): 637–41.
13. Player MS, Gill JM, Mainous AG 3rd, *et al*. An electronic medical record-based intervention to improve quality of care for gastro-esophageal reflux disease (GERD) and atypical presentations of GERD. *Qual Prim Care* 2010; 18 (4): 223–9.



14. DeShazo JP, Hoffman MA. A comparison of a multistate inpatient EHR database to the HCUP Nationwide Inpatient Sample. *BMC Health Serv Res* 2015; 15 (1): 384.
15. Shafiq A, Goyal A, Jones PG, *et al.* Serum magnesium levels and in-hospital mortality in acute myocardial infarction. *J Am Coll Cardiol* 2017; 69 (22): 2771–2.
16. Goyal A, Spertus JA, Gosch K, *et al.* Serum potassium levels and mortality in acute myocardial infarction. *JAMA* 2012; 307 (2): 157–64.
17. Bath J, Smith JB, Kruse RL, *et al.* Cohort study of risk factors for 30-day readmission after abdominal aortic aneurysm repair. *Vasa* 2019; 48 (3): 251–61.
18. Kruse RL, Hays HD, Madsen RW, *et al.* Risk factors for all-cause hospital readmission within 30 days of hospital discharge. *J Clin Outcomes Measures* 2013; 20 (5): 203–14.
19. R Core Team. R: *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018. <http://www.R-project.org/>.
20. Fortin Y, Crispo JA, Cohen D, *et al.* External validation and comparison of two variants of the Elixhauser comorbidity measures for all-cause mortality. *PLoS One* 2017; 12 (3): e0174379.
21. Upadhyay DK, Sittig DF, Singh H. Ebola US Patient Zero: lessons on misdiagnosis and effective use of electronic health records. *Diagnosis (Berl)* 2014; 1 (4): 283–7.
22. Blumenthal D, Tavenner M. The “meaningful use” regulation for electronic health records. *N Engl J Med* 2010; 363 (6): 501–4.
23. Lex A, Gehlenborg N, Strobel H, *et al.* UpSet: visualization of intersecting sets. *IEEE Trans Vis Comput Graph* 2014; 20 (12): 1983–92.
24. Stewart CC, Lu CY, Yoon TK, *et al.* Impact of ICD-10-CM transition on mental health diagnoses recording. *eGEMS* 2019; 7 (1): 14.
25. Slavova S, Costich JF, Luu H, *et al.* Interrupted time series design to evaluate the effect of the ICD-9-CM to ICD-10-CM coding transition on injury hospitalization trends. *Inj Epidemiol* 2018; 5 (1): 36.
26. Carroll NM, Ritzwoller DP, Banegas MP, *et al.* Performance of cancer recurrence algorithms after coding scheme switch from International Classification of Diseases 9th Revision to International Classification of Diseases 10th Revision. *JCO Clin Cancer Inform* 2019; 3: 1–9.
27. Panozzo CA, Woodworth TS, Welch EC, *et al.* Early impact of the ICD-10-CM transition on selected health outcomes in 13 electronic health care databases in the United States. *Pharmacoepidemiol Drug Saf* 2018; 27 (8): 839–47.