IBM Watson Health™

WHITE PAPER

Interoperability in the Age of Population Health Management – Part 1

Contents

Population health management (PHM) challenges healthcare organizations to leverage the vast array of data they have in unprecedented ways. Key executives within those healthcare organizations are also being asked to use new information technology tools to manage populations.

Chief financial officer - needs to quickly assess the financial performance of value-based payment contracts and other alternative payment models.

Chief medical officer oversees clinical quality performance and patient satisfaction; needs proactive information to guide process improvement.

Frontline clinicians - need curated data and analytics to manage population health.

Physicians - need to know what other care a patient may be receiving outside the healthcare system and must send reminders and alerts to their patients between visits.

Care managers - coordinate the care of high-risk patients; need to understand and address their social and other determinants of health to encourage patients in self-care and understand their unique risk factors.

Individual patients - need tools that will help them better engage with their care teams and comply with their care plans.

Today, most healthcare organizations lack access to data from across the continuum of care, because their electronic health records (EHRs) are not interoperable with those of other providers and healthcare organizations. Though some EHR vendors have committed to addressing this issue, today the information in these disparate EHRs can't be exchanged in a form that clinicians can review and act upon. While progress has been made in exchanging standardized care summaries between different EHRs, these summaries are in the form of documents, not discrete data.

Also at an early stage is the ability of EHRs to integrate with outside applications designed for population health management. As with EHR-to-EHR data exchange, the proprietary code underlying EHRs prevents them from linking to these PHM applications without special interfaces. Standardized messaging formats can partly overcome this barrier, but it will persist until the industry realizes the potential of a new technology known as Fast Healthcare Interoperability Resources (FHIR). Expected to come into general use in the next year or two, FHIR will allow EHR users to expose their clinical content to whatever applications they authorize to expand their EHR's capabilities.

Meanwhile, healthcare organizations (HCOs) must address two primary challenges in building a data infrastructure for PHM:

1. The ability to integrate and analyze clinical, administrative and claims data from multiple sources in near real-time.
2. The ability to insert the relevant integrated data and analytic insights into the workflows of the care team members so they can use the information in patient care.

This two-part series explains how both of these goals can be accomplished with today's technology. Part two also describes the standards framework called Fast Health Interoperability Resources (FHIR) and shows how it will facilitate the integration of outside PHM applications with EHRs to help support population health management.

Challenge #1: Interoperability

The Office of the National Coordinator for Health IT (ONC) defines interoperability as "the ability of a system to exchange electronic health information from other systems without special effort on the part of the user."[1] This has long been a key goal of healthcare leaders and policy experts, who say that interoperability is a prerequisite for obtaining true value from IT systems. In its 2017 interoperability roadmap, however, ONC acknowledges that interoperability has not yet arrived.[2] A recent survey of healthcare executives, similarly, found only 8% of providers had interoperability with other providers and were integrating outside data into their workflow.[3]

Consequently, most of the health information exchange that occurs today is confined to sending and receiving documents such as care summaries. These documents are generally transmitted in the Consolidated Clinical Data Architecture (CCDA) format. Direct secure messaging, which can be used to convey these care summaries, is growing but is still not widespread. Record locators are increasingly being used to retrieve documents from distributed databases.

Some health information exchanges (HIEs) have achieved partial success in their markets, but only a quarter of public HIEs consider themselves to be financially stable.[4] Even if the exchanges did find a way forward, most HIEs provide hospital-derived data to a community, rather than a bidirectional data flow between inpatient and outpatient sources. This approach does not align well with PHM, "which requires a more fluid and interactive data flow across a range of care settings," noted a recent report from Chilmark Research. In addition, HIEs collect only a small portion of the data needed for PHM, Chilmark observed.[5]

The vast majority of HCOs now use enterprise data warehouses (EDWs) to aggregate and analyze their own data. But their ability to do so is largely dependent on their size. In a recent survey, 94% of hospitals with 200 or fewer beds said they were not adequately capturing the data they needed for population health management analytics.[6]

One reason may be the lack of interoperability between the systems of hospitals and the ambulatory-care clinics they own or partner with. Also, it is difficult for traditional EDWs to produce reports that have not been preprogrammed.

The data lake: an alternative model

However, there is an alternative. A PHM IT infrastructure can be based on a "data lake," an advanced type of EDW that uses an ad hoc approach to gather data and run reports. In contrast to the relational databases used in most enterprise data warehouses, the data lake employs frameworks such as Hadoop, designed for distributed storage and distributed processing of large datasets in cloud-based computer clusters.

Among the advantages of this big data approach are flexibility, timeliness, and scalability. The data lake is flexible because rules can be quickly written when a new type of report is needed. As a result, these new reports can be generated in days or weeks rather than the months it typically takes to rewrite preformatted reports.

Because the process of mapping the data to standard code sets and terms is semi-automated, reports can be run and registries updated much faster than in the traditional data warehouse. As a result, a healthcare system can deliver curated, analyzed data from multiple sources to its providers, care coordinators, and PHM managers within 24 hours after the original data was entered in an EHR.

An effective IT infrastructure for PHM should be able to integrate and normalize clinical data from many different sources in near real time. Without that capability, it is difficult for providers and care managers to find out what has been done for patients outside of their organization. That knowledge is essential for care coordination and the reduction of waste in healthcare.

IBM Watson Health has adopted the data lake platform as its underlying architecture for integrating the clinical and claims data that are needed for population health management. The company believes that this approach can replace health information exchanges as a conduit for the interchange of critical data, while allowing the information to be aggregated, normalized and analyzed quickly enough to be useful at the point of care.

Data integration will only increase in importance as new kinds of data become available. Patient-generated data, data from post-acute care providers, behavioral health data, genetic data, and data related to social determinants of health will all eventually find their place in PHM. HCOs should consider how best to integrate a wide range of data that will enable their organizations to thrive under value-based reimbursement.

IBM Watson Health, through the acquisition of Truven, has access to claims data from 210 million health plan members, including those in Medicare and Medicaid managed care plans. Truven and Watson Health can combine claims, clinical, and public health data and integrate them in an enhanced CCDA that can be consumed by any certified EHR. That information can also be used to identify care gaps and create a coordinated care management plan for the patient.

Challenge #2: Data mapping and cleanup

Clinical data includes information from multiple EHRs, other healthcare systems, and the systems of post-acute-care (PAC) providers, laboratories, imaging centers and pharmacies, among other entities. All of this data must be aggregated, normalized and integrated before analytics can be applied to it.

A database designed for PHM must include administrative data, such as billing, scheduling and demographic data. The billing data includes diagnostic and procedural codes that are critical to population health management solutions. An analytic program, for example, uses scheduling data to determine when a patient last visited his or her doctor—a key data point for detecting care gaps. And the demographic data includes information on patient identity, age and gender.

Claims data, which provides a very broad view of a patient's care, can also be integrated with clinical data. But it requires considerable technical expertise to parse claims information, and great care must be taken when combining it with clinical data. For example, when a provider orders a test to rule out a diagnosis, that doesn't mean that the patient has that diagnosis; yet a code for it may appear on a claim for the test. Claims must be compared with clinical data to eliminate these kinds of errors.

Data collection

The first step in data collection is to identify the data sources that are available and necessary to meet an HCO's goals. Once that has been done, the HCO's data vendor can build interfaces with those data sources. An integration engine can then pull the information from these sources into the data lake, where each bit of it receives an XML metadata tag that identifies that datum uniquely. It will always remain in its native state in the data lake, but its tag will link it to its curated version in a standardized database.

Let's assume that this unified data model is based on SNOMED-CT, the broadest compendium of medical terms. A map using that ontology would convert ICD-9 and ICD-10 diagnostic codes and CPT procedure codes into SNOMED codes.

To map pharmacy claims, medication orders, and drug administration records, the integration engine would use a combination of SNOMED and RxNorm, which translates messages between drug-related systems that don't use the same software and vocabulary.

Laboratory data poses a greater challenge, because each lab has its own terminology and coding standards. The Regenstrief Institute has created an open map that can be used to translate these varying codes into the Logical Observation Identifiers Names and Codes (LOINC) hierarchy. However, where no logical target exists in LOINC for an observation, an auxiliary map must be used to index and search for it.

Mapping challenges

Most diagnostic data and at least half of procedural data are already coded in a standard format when they come into the data lake. That should make it relatively easy to map this kind of information to SNOMED-CT. However, many hospitals have customized these codes for their own purposes. When that occurs, the mapping algorithm doesn't work and manual effort is needed to map this data to SNOMED.

Early in the mapping process, some data may be excluded. This excluded data may include invalid dates or technically or clinically implausible values, for example.

Data may be missing from the discrete fields in EHRs. Much of this data is trapped in free text, as we explain later. But in some cases, the missing data can be imputed from the existing data by applying certain rules. Among the data points that may be imputed are body mass index (BMI), body surface area, clinical diagnoses, length of stay, medication end dates and sig, units of measure, and death date.

For example, an EHR typically generates a BMI only if a patient's height and weight are measured at the same time. But a programmed rule can specify that if one value is documented without the other, the closest measurement of the latter value in time can be used. So if a patient comes in and has her weight but not her height measured, her BMI can be imputed using the height from her last annual exam. Another example: If a patient has an HbA1c of 9.1 and insulin has been ordered for daily administration, then the diagnosis of diabetes can be imputed.

Semantic interoperability

To be reliable and effective, the integration engine, with the help of the mapping staff, must be able to map 90%-95% of the source data to a standard ontology. On the clinical side, however, much of the source data may not be standardized. Typical data that may not be in a standardized format includes lab results, allergies, immunizations, and health maintenance histories. With the exception of lab results, the other data types may have custom workflows that vary from one EHR to another.

In addition, roughly 80% of the data in an EHR is unstructured, including free text, documents, and images.[7] Medical observations may be dictated by doctors and found in transcribed notes or in documents received from outside sources, such as consultation reports and discharge summaries.

It is too time-consuming and expensive to go through every single document and reconcile the different terms and abbreviations used for the same medical concepts. Methods exist to automate this process as much as possible. What these software applications seek to accomplish is called "semantic interoperability." This is a form of mapping that links the original terms or abbreviations to a standard vocabulary that is associated with standard codes.

To illustrate how difficult this is, the participants in one accountable care organization spelled or denoted the word "bypass," as in "coronary artery bypass grafting," in 243 different ways.[8] There are 100 different terms and codes identified for "high blood pressure." And there are 70 to 80 distinct healthcare terminology vocabularies, ranging from billing codes to clinical vocabularies created by physicians, nurses, hospitals, pharmacists, and laboratories.[9]

> The latest semantic interoperability tools of IBM Watson Health can automate the process of matching about 80% of terms and codes to the standard model. Beyond that, manual review is required to raise the matching ratio above 90%.[10]

Natural language processing

Natural language processing (NLP) is the ability of computers to understand human language and written text. Cognitive computing systems, which include a form of NLP, can extract insights from unstructured data.

For example, NLP has been used to support quality reporting by some physician groups. One of the quality measures they selected required a numerical value related to the left ventricular ejection fraction in heart failure patients. Because this value is not usually stored in its own EHR field, healthcare systems have to hire nurse abstractors to review charts and enter the relevant values in the quality reports. However, a cognitive computing system with NLP, which has been trained to read echocardiogram reports, can recognize that discrete data value and input it into the reports.

NLP is also being used with lab results that include qualitative information in text form. The NLP standardizes that text to categories such as positive, negative, normal and abnormal.

Mapping maintenance

As mentioned earlier, HCOs frequently customize their EHRs to meet their needs and goals. As a result, the data mapping performed on their behalf can quickly become outdated. Naturally, this is much more complicated when a number of different data sources are involved. Moreover, the public and private payers that seek quality data from healthcare providers often change their measures, and HCOs may pick different metrics from one year to the next. When this happens, different kinds of data must be integrated to produce quality reports.

To ensure that these changes do not reduce the accuracy of data mapping, the data must be remapped at regular intervals or as needed. How often this should be done may depend on how many alerts are received in audit logs. These alerts are fired when new codes or terms that the system hasn't seen before arrive in the data lake.

Conclusion

The data lake approach to data integration represents a quantum leap beyond the static, inflexible approach of traditional data warehouses. Combined with robust, semi-automated data mapping methods, the data lake enables healthcare organizations to quickly deliver custom reports and insights based on comprehensive patient data and the latest studies in the literature.

Semantic interoperability is also required to match terms that are not coded to a standard vocabulary. And natural language processing can be used to extract insights from unstructured data, which forms about 80% of an EHR's content.

In part two of this series about interoperability, we look at the critical step that follows data integration: EHR workflow integration.

References

[1] Office of the National Coordinator for Health IT, "Connecting Health and Care for The Nation: A Shared Nationwide Interoperability Roadmap," final version, 2015, ix.

[2] Office of the National Coordinator for Health IT, "Connecting Health and Care for The Nation: A Shared Nationwide Interoperability Roadmap," draft version, 2015, 10.

[3] Marla Durben Hirsch, "KLAS: Only 6 percent of providers hit interoperability 'home runs,'" Fierce Healthcare, Oct. 19, 2016, http://www.fiercehealthcare.com/it/klas-only-6-providers-had-interoperability-home-run?utm_medium=nl&utm_source=internal&mrkid=629433&mkt_tok=eyJpIjoiWWpRME5qWmtNakpsTWpNeiIsInQiOiJ0ZDNzbmdpQ3BIczVwV0hNYjVTMTRIZ2V2UytoRXJLYTFMQlwvZUZQb09XekZtSEhhVjVOb2Z3em90aFJcL0hLK1JHUmt1c25XaEdOcmNOcHNJNVZqQVJiUmFZMkJ5OHBERFNMXC9wRGtIWkluOD0ifQ%3D%3D

[4] Robert S. Rudin, Aneesa Motala, Caroline Lubrick-Goldzweig, and Paul G. Shekele, "Usage and Effect of Health Information Exchange: A Systematic Review," Annals of Internal Medicine, v.161, no. 11, Dec. 2014, pp. 803-812.

[5] Chilmark Research, "Population Health Management: Aligning IT Solutions to Strategy," April 2015, 25.

[6] Black Book Market Research, "9 Healthcare Tech Trends in 'The New Year of Uncertainty," press release, Dec. 20, 2016, https://blackbookmarketresearch.newswire.com/news/9-healthcare-tech-trends-in-the-new-year-of-uncertainty-black-book-18314838

[7] John Smithwick, "Unlocking the value of unstructured patient data," Becker's Health IT and CIO Review," Aug. 25, 2015, http://www.beckershospitalreview.com/healthcare-information-technology/unlocking-the-value-of-unstructured-patient-data.html

[8] IBM Watson Health Care, internal communication.

[9] George Schwend, president and CEO of Health Language Inc., personal communication.

[10] IBM Watson Healthcare, internal communication.

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America July 2017

Note: Statistics and testing provided by the Explorys Analytics Team and Definitive Healthcare, which provides powerful intelligence on hospitals, physicians, and healthcare providers; http://www.definitivehc.com/.

PROV 18119 0717

IBM **Watson Health**™