

# Big Data Applications In Population Health Management

Tyler Peterson

Indiana University - School of Informatics, Computing, and Engineering

711 N. Park Avenue

Bloomington, Indiana 47408

typeter@iu.edu

## ABSTRACT

Healthcare providers are experiencing pressure to reduce costs while also delivering increasingly high quality care. External forces in the form of alternative reimbursement models and programs offering both incentives and penalties have spurred healthcare organizations to find opportunities for increasing the value of their work. Given the complexity of the healthcare system, the opportunities are vast in number. Preventing hospital admissions, management of chronic conditions and the early detection of potentially deadly conditions are a few of the major initiatives. The shared attribute of these three opportunities is that the solutions are often most effective when directed towards people who are going about their day-to-day lives in the community, as opposed to those who are currently confined to hospital beds or in an exam room. Through these programs, providers are compelled to proactively reach out to all people who make up the population they serve. To understand the needs of a population, healthcare providers must embrace big data. Novel analysis and presentation of massive, heterogeneous datasets is essential. Big data applications and analytical tools are appearing at the bed side, in exam rooms and on patients themselves as providers seek to harness the power of big data.

## KEYWORDS

i523, hid331, Population Health, Predictive Analytics, Electronic Medical Records, EMRs

## 1 INTRODUCTION

The United States spent \$3.2 trillion on healthcare in 2015, 5.8 percent higher than the previous year[8]. This amounts to nearly \$10,000 per person living in America[8]. Chronic diseases such as diabetes, cancer and cardiovascular disease contribute to 70 percent of US deaths and incur 75 percent of healthcare expenditures[9]. Despite these high costs, the US lags behind other countries in quality[7]. In other words, we are paying more for less. Healthcare providers are looking to use big data to aid their efforts of reducing costs and increasing the quality of care. The system has a massive amount of data at its disposal, and the rate at which it is accumulating is increasing. In 2011, the US health system possessed 150 exabytes, and the total will soon be counted in zettabytes[14].

The proposed solutions for achieving less expensive, higher quality care are numerous and varied. Several proposals seek to alter the behavior of healthcare providers that have grown accustomed to the fee-for-service reimbursement model. Traditionally, providers are reimbursed for services rendered. Redundant tests, unnecessarily readmitted patients, and frequent emergency room visits generate a profit. This is a financial arrangement that rewards volume without consideration to value[11]. It directs attention to patients who can

be immediately provided with services, namely patients who are either in the hospital or at an outpatient appointment. This fee-for-service reimbursement model does not incentivize healthcare providers to look beyond its immediate customers and take the initiative to provide services to the community at-large.

Alternative payment models are engineered with the intention of changing that behavior. The Medicare Shared Savings Program (MSSP) is a type of Accountable Care Organization (ACO) that provides a framework for healthcare organizations to have a level of accountability for the quality, cost and patient experience of an assigned population. MSSP participants choose one of four financial risk arrangements. One track lets providers avoid any penalty, or risk, in the case that they do not lower their Medicare expenditure growth. The other three tracks offer increasingly higher risks and rewards[4]. In order to receive a share of any cost savings, ACOs must demonstrate the delivery of high quality care by reporting the organization's performance on quality measures that can be categorized into four domains - patient experience, care coordination, preventative health and at-risk population management[2].

Other initiatives penalize providers for delivering sub-standard care. In 2016, the US government penalized 2,597 hospitals for excessive 30-day hospital readmissions[15]. If patients are initially admitted with a heart attack, heart failure, pneumonia, chronic lung disease or for a hip/knee replacement procedure and are subsequently readmitted to a hospital within a month of the initial stay, a hospital is considered responsible. The penalties for those readmissions amounted to \$528 million nationally in 2016, \$108 million higher than 2015[15].

Entire communities stand to benefit from these programs. Providers must proactively engage with patients who are going about their day-to-day lives while also providing high quality care within their hospitals and at their clinics. To understand how to approach each individual that makes up their community, healthcare providers must harness the power of big data. Mitigating a hospital admission before it occurs or discovering and treating a condition before it worsens requires methodical data collection, pinpointed data analysis, and deliberate, compassionate execution of proactive healthcare delivery. Big data applications and analytical tools are essential for accomplishing those demands.

## 2 BIG DATA IN POPULATION HEALTH

Big data applications and analytics are well-suited for approaching the issues and opportunities described above because while health care is often described in macro terms, the meaningful interactions happen at the micro level. High-level, aggregated datasets may describe a population, but don't provide the necessary depth for understanding how one patient's needs and circumstances are unique from all the rest. There are various types of data that help illustrate

a patient's overall health picture. Medical claim data includes diagnoses, procedures, dates, cost and points of care. Electronic medical records (EMRs) also includes diagnoses, procedures and dates, while also cataloging lab values, free-text notes, images and medication lists. Clinical trial data, patient satisfaction survey data, genomics data and medical device data also contributes to the massive amount of healthcare data available to providers[10].

This data can be used to meet the nuanced demands of alternative reimbursement models and avoid readmission-related penalties. These programs have attributes that lends themselves to big data applications. MSSP participants need to attest to 31 measures in 2017. These measures, for example, address diabetics with poor hemoglobin A1c control, all-cause, unplanned admissions for heart failure patients, use of imaging studies for low back pain, and patients' perceived quality of communication with providers[3]. Data is essential for identifying patients who fall within the scope of each metric, determining which patients have already met the goal of the measure, and engineering processes to help make providers aware of the patients who have yet to receive the recommended intervention. For example, a healthcare organization must identify their diabetic patients (typically with ICD-10 diagnosis codes), determine which of those patients had their hemoglobin A1c tested within the measurement period, and had a lab result within the accepted value range. Patients may fail the measure in one of two ways: a patient either has not been tested within the measurement period or the patient has a lab value outside of the acceptable range. Patients in the former category should be contacted by the provider and have an appointment scheduled to have the lab drawn. The patients in the latter category should be treated in a manner that brings the hemoglobin A1c within the acceptable range and followed closely by a care team.

To avoid penalties associated with readmissions, some healthcare providers are employing advanced techniques, such as machine learning, to identify patients who are at high risk of being readmitted within 30 days of the initial hospital stay. Mount Sinai Health System in New York, NY, developed a predictive model to evaluate heart failure patients for risk of readmission. The model building began by analyzing 4,205 attributes, including 1,763 diagnosis codes, 1,028 medications, 846 laboratory measurements, 564 surgical procedures, and 4 types of vital signs[13]. Mount Sinai concluded that their model featured in the research study outperformed the previous models used to assess their heart failure patients, while conceding that the model needs to be updated and recalibrated with several years of data from several different hospital sites[13]. In other words, even more data is needed. In the meantime, this model can still be used to analyze each heart failure patient prior to discharge for the likelihood of readmission. Care teams can then dedicate extra resources to the especially high-risk patients.

Wearable technology has also infiltrated the healthcare space, especially devices that can remotely and wirelessly monitor patients' vitals and symptoms. The data feeds can be used by providers to assess the effectiveness of (and adherence to) medications, observe lifestyle habits, or recommend that a patient schedule a follow-up appointment or go to an emergency room[1]. There are hundreds of thousands of mobile health apps available in app stores, and more than half of these are geared for patients with chronic diseases[1]. This technology has appeared in clinical trials as well. A study

determined that patients with type 2 diabetes who monitored blood glucose with an app achieved greater reduction in hemoglobin A1c results compared to patients who did not use an app[6].

### 3 INFRASTRUCTURE

The infrastructure needed to support these efforts is complex. A cornerstone of enabling big data analysis in healthcare is EMR software. EMRs replace the paper chart as the location for all details related to patient care. These information systems gather a wide variety of information, including patient encounters, lab results, medications, diagnoses, and procedures, as well as demographic and socioeconomic information, among many other data elements. Providers may also add notes by typing or through dictation software. The information is stored in data warehouses that can be queried and analyzed in a way unimaginable in the era of paper charts. EMRs can also be programmed to remind or notify a provider that, for example, a patient meets the criteria for the MSSP colonoscopy screening measure and has not had a colonoscopy in ten years, so an appointment for the procedure should be scheduled. As of May 2016, 96 percent of non-Federal acute care hospitals had adopted a certified EMR[5].

Health care data is growing in such a way that it benefits from big data applications such as Hadoop and MapReduce, which create a framework capable of handling massive amounts of structured data, such as discrete lab values and diagnosis codes, and unstructured data, such as physician notes. Hadoop breaks the large datasets into smaller subsets, MapReduce processes those subsets independently and in parallel, and the processed subsets are combined into a final result[12].

Data visualization tools are also essential for communicating key messages in data. Tools such as Tableau and Qlikview, and open source code libraries such as Plotly and Bokeh (written for Python), allow savvy users to present large, complex data sets in visually compelling ways to quickly communicate important ideas. Dashboards can promote exploratory data analysis, and can engage even those who are not technical through easy to use point-and-click user interfaces.

### 4 CONCLUSION

Big data applications are capable of turning data into insights, and this is critical for aiding healthcare providers in their efforts to evolve the way they practice medicine. EMRs will continue to amass vast amounts of information about patients and their unique characteristics and needs. Programs and policies will continue to foster the mindset that healthcare providers must actively consider all individuals who constitute the population they serve, not just the patients actively in a hospital or present in a clinic. Big data applications will continue to be engineered to deliver the right information to the right provider. These tools will promote the most beneficial action for each individual patient.

### ACKNOWLEDGMENTS

The author would like to thank Professor Gregor von Laszewski and his teaching assistant for their help with Github, Latex and JabRef.

## REFERENCES

- [1] Linda Brookes. 2017. Can Technology Transform Chronic Disease Management? (April 2017). <https://www.medicalnewstoday.com/articles/317016.php>
- [2] Centers for Medicare and Medicaid Services. 2017. Accountable Care Organization 2017 Quality Measure Narrative Specifications. (January 2017). <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/2017-Reporting-Year-Narrative-Specifications.pdf>
- [3] Center for Medicare and Medicaid Services. 2017. Medicare Share Savings Program Quality Measure Benchmarks for 2016 and 2017 Reporting Years. (December 2017). <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/MSSP-QM-Benchmarks-2016.pdf>
- [4] Centers for Medicare and Medicaid Services. 2017. New Accountable Care Organization Model Opportunity: Medicare ACO Track 1+ Model. (July 2017). <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/New-Accountable-Care-Organization-Model-Opportunity-Fact-Sheet.pdf>
- [5] JaWanna Henry, Yuri Pylypchuk, Talisha Searcy, and Vaishali Patel. 2016. Adoption of Electronic Health Record Systems Among U.S. Non-Federal Acute Care Hospitals: 2008-2015. (May 2016). <https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php>
- [6] Can Hou, Ben Carter, Jonathan Hewitt, Trevor Francisa, and Sharon Mayor. 2016. Do Mobile Phone Applications Improve Glycemic Control (HbA1c) in the Self-management of Diabetes? A Systematic Review, Meta-analysis, and GRADE of 14 Randomized Trials. *Diabetes Care* 39, 11 (November 2016), 2089–2095. <http://care.diabetesjournals.org/content/39/11/2089>
- [7] Rabah Kamal and Cynthia Cox. 2017. U.S. Health System is Performing Better, Though Still Lagging Behind Other Countries. (May 2017). <https://www.healthsystemtracker.org/brief/u-s-health-system-performing-better-though-still-lagging-behind-countries/#item-start>
- [8] Anne B. Martin, Micah Hartman, Benjamin Washington, Aaron Catlin, and National Health Expenditure Accounts Team. 2017. National Health Spending: Faster Growth In 2015 As Coverage Expands And Utilization Increases. *Health Affairs* 36, No. 1 (2017), 166–176. <http://content.healthaffairs.org/content/36/1/166>
- [9] Farshad Fani Marvasti and Randall S. Stafford. 2012. From Sick Care to Health Care - Reengineering Prevention Into The U.S. System. *The New England Journal of Medicine* 367, 10 (September 2012), 889–891. <http://www.nejm.org/doi/full/10.1056/NEJMp1206230#t=article>
- [10] Carol McDonald. 2016. How Big Data Is Reducing Costs And Improving Outcomes In Health Care. (June 2016). <https://mapr.com/blog/reduce-costs-and-improve-health-care-with-big-data/>
- [11] Harold D. Miller. 2009. From Volume To Value: Better Ways To Pay For Health Care. *Health Affairs* 28, 5 (2009), 1418–1428. <http://content.healthaffairs.org/content/28/5/1418.abstract>
- [12] Elizabeth O'Dowd. 2016. How Hadoop Supports Healthcare Data Analytics Infrastructure. (October 2016). <https://hitinfrastructure.com/news/how-hadoop-supports-healthcare-data-analytics-infrastructure>
- [13] Pacific Symposium on Biocomputing (Ed.). 2016. *Predictive Modeling of Hospital Readmission Rates Using Electronic Medical Record-Wide Machine Learning: A Case-Study Using Mount Sinai Heart Failure Cohort*. Vol. 22. Pacific Symposium on Biocomputing. <https://www.ncbi.nlm.nih.gov/pubmed/27896982>
- [14] Wullianallur Raghupathi and Viju Raghupathi. 2014. Big Data Analytics In Healthcare: Promise And Potential. *Health Information Science and Systems* 2, 3 (2014). <https://www.biomedcentral.com/track/pdf/10.1186/2047-2501-2-3?site=hissjournal.biomedcentral.com>
- [15] Jordan Rau. 2016. Medicare's Readmission Penalties Hit New High. (August 2016). <https://khn.org/news/more-than-half-of-hospitals-to-be-penalized-for-excess-readmissions/amp/>

## 5 BIBTEX ISSUES

Warning—entry type for "editor09" isn't style-file defined

—line 5 of file report.bib

Warning—entry type for "editor15" isn't style-file defined

—line 17 of file report.bib

Warning—entry type for "editor06" isn't style-file defined

—line 45 of file report.bib

Warning—entry type for "editor14" isn't style-file defined

—line 87 of file report.bib

Warning—entry type for "editor03" isn't style-file defined

—line 99 of file report.bib

Warning—entry type for "editor04" isn't style-file defined

—line 111 of file report.bib

Warning—entry type for "editor07" isn't style-file defined

—line 123 of file report.bib

Warning—entry type for "editor13" isn't style-file defined

—line 149 of file report.bib

Warning—entry type for "editor05" isn't style-file defined

—line 174 of file report.bib

Warning—can't use both volume and number fields in editor08

Warning—page numbers missing in both pages and numpages fields in editor12

(There were 11 warnings)

## 6 ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

### 6.1 Writing Errors

Errors in title, e.g. capitalization - in