2023

The Impact of Big Data Utilization on Quality Improvement in Inpatient Facilities

Lakyn Hare
LakynHare510@outlook.com

Lakyn Hare
12/02/2023
HCA 695

THE IMPACT OF BIG DATA UTILIZATION ON QUALITY IMPROVEMENT IN INPATIENT FACILITIES

ABSTRACT

Introduction: Poor quality in healthcare has resulted in avoidable patient complications, including readmission rates. Big data in healthcare can be analyzed and built into a tools, with machine learning, to aid in reduced readmission rates and overall positive patient outcomes.
Purpose of the Study: The intention of this study was to evaluate the ways that big data can be analyzed to improve healthcare, specifically readmissions, patient outcomes, and show cost savings. This study examined different ways that big data could be used in concordance with machine learning, including predictive analysis, to make these improvements.
Methodology: The hypothesis was the use of big data has improved quality categories in inpatient facilities led to improved patient outcomes, decreased readmission rates, and reduced costs. 27 publications were included and limited to the English language and were published between the year(s) of 2010 and 2023.
Results: The study displayed many solutions to different concerns within the scope of readmissions and patient outcomes through machine learning tools alongside big data. Examined were solutions for reducing readmission rates, and patient outcomes which further included addressing appointment no-shows, disease risks, and outcome outlook.
Discussion: The hypothesis of this study was partially conclusive, as improvement to readmission rates and patient outcomes were examined, but financial data was unavailable to confirm cost savings potential. The study limitations included the inability to obtain financial statistics and the potential for machine learning training data to not be clean. An interview with

an expert in quality was also conducted and subject opinions were displayed alongside its corresponding subcategory.

Conclusion: The research provided descriptive and valuable data on the abilities of big data and machine learning techniques. Given the study limitations, there has been shown a need for research and data on the common use of these techniques within inpatient facilities.

Keywords: Big Data, Big Healthcare Data, Costs, Patient Outcomes, Quality Improvement

INTRODUCTION

Poor quality in healthcare has been a global issue, but the United States (US) has continued to fall behind in multiple areas of quality (WHO, 2018; ARHQ, 2018). These areas of poor quality have included misuse, overuse, underuse, and variation in the use of services, which has resulted in avoidable issues, such as additional complications for patients, misdiagnosis, injury and death, and excess costs (ARQH, 2018). One of these categories alone, overtreatment, was responsible for 2%-2.7% of healthcare spending in the U.S. in 2019. These unnecessary services added an estimated $76 billion to the U.S. healthcare spending in 2019. Also in 2019, failure of care delivery cost the U.S. an estimated $102-166 billion, and failure of care coordination cost roughly $27-78 billion (Health Affairs Research Brief, 2022; Shrank et al., 2019).

In addition to quality, patient outcomes, and costs, readmission rates have also been a concern. Hospital readmission rates have been an important quality indicator as it has had increased use as an indicator for quality improvement and cost control (Fischer et al., 2014). The Affordable Care Act (ACA) reform made reducing hospital readmission rates a priority when in 2011 there were 3.3 million hospital readmissions, or $41.3 billion in associated costs (NEJM Catalyst, 2018b). Medicare created the Hospital Readmissions Reduction Program (HRRP), which created penalties for acute-care hospitals that have higher 30-day readmission rates relative to other facilities, and this implementation has shown declined rates (NEJM Catalyst, 2018b). With the implementation of HRRP, readmission rates fell to 17.8% in 2015 from 21.5% in 2007 (Zuckerman et al., 2016). The Center for Medicare and Medicaid Services (CMS) has also been improving quality by focusing on multiple areas such as adopting measures that have been patient-centered, outcome-based, and high-impact, as well as minimizing provider burden, and using alternative payment models (CMS, 2021).

The use of big healthcare data can positively affect patient outcomes, healthcare costs, and readmission rates through several conceptual approaches and frameworks such as predictive analysis and machine learning (Kumar et al., 2018). Big data has a vast amount of information and data on a specific topic. Big Data in healthcare has been sourced from, but not limited to, medical imaging, Electronic Health Records (EHRs), payor records, genomic sequencing, pharmaceutical research, wearables, and medical devices (NEJM Catalyst, 2018a). Big data differs from traditional health data used for decision-making by three characteristics, referred to as the 3-V's of Big Data which are high volume, high velocity, and high variable (NEJM Catalyst, 2018a). The data is available at a very high volume, and it moves at a high velocity and spans the health industry's massive digital universe as it derives from many sources, and has provided a highly variable structure and nature (NEJM Catalyst, 2018a).

Big data could be analyzed for providers and administrators to make informed decisions for a multitude of treatments and services (Tulane University, 2021). With current available technology, accessing and analyzing data has never been easier and it has suggested that patient outcomes could be improved by big data through resources like Electronic Health Records (EHR), testing machines, and exam results (Tulane University, 2021). Although there has been potential for the use of big data in healthcare, deciding on the allowable uses of the data could be challenging when also trying to preserve patient privacy and security (Abouelmehdi et al., 2018).

Analyzing big data could lead to many opportunities for cost reduction in healthcare through predictive systems and infrastructure; cost reduction opportunities include high-cost patients, readmissions, triage, decompensation, adverse events, and treatment optimization (Bates, 2014).

The purpose of this research was to analyze the effects of big data being used to improve quality, specifically readmission rates, and reduce associated costs in inpatient facilities.

METHODOLOGY

The hypothesis was the use of big data has improved quality categories in inpatient facilities led to improved patient outcomes, decreased readmission rates, and reduced costs.

The methodology for this qualitative study was a literature review and a semi-structured interview with an expert in quality improvement in healthcare. The literature reviewed and cited was peer-reviewed, research articles and government websites, found by Marshall University's database finder, ProQuest, PubMed, and Google Scholar. Google was also used to find government websites, articles, and data. Keywords used in the search included, "big data" OR "big healthcare data", AND "quality improvement", OR "readmission rates", OR "costs", OR "patient outcomes".

The 27 publications that were reviewed and used were limited to the English language and were published between the year(s) of 2010 and 2023. Literature was selected based on whether the abstract included relevant information and data to quality improvement efforts through the use of big healthcare data, specifically decreased readmission rates and reduced costs. The information searched for was specific to the effects that big data in healthcare had on patient outcomes, such as reducing readmission rates.

A semi-structured interview with an expert in healthcare quality was completed virtually using Microsoft Teams and was recorded and transcribed. Following transcription, the recording was permanently deleted. Consent was obtained from the subject verbally prior to the questionnaire. The list of questions asked to the subject can be found in the references section. The plan for this study was reviewed by the Marshall University Institutional Review Board (IRB) and was deemed exempt on October 12, 2023.

This study's conceptual framework (Figure 1) was adapted from the research framework of Yao, Chu, and Li (Yao et al., 2010). The framework displays the reasoning of and approach to big data and machine learning for improving healthcare quality, specifically readmission rates and patient outcomes. The adoption of the use of big data begins with the need for tools and algorithms to support clinical and administrative functions due to big data's ability to support these improvements.

The preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) method was used to identify the relevant citations in this research (PRISMA, 2021). Using the PRISMA method, a total of 122 records were identified. References were only included (N=27) if they had useful information to this research and hypothesis, others were excluded (N=95) as they did not meet inclusion principles. Figure 2 displays the visual process of the PRISMA method. Fully reviewed references (N=11) were used in total and are displayed in the references section of this paper. The search was executed by LH and verified by AC, who served as a secondary reviewer and reviewed references met inclusion criteria.

RESULTS

Big Data for Reduced Readmission Rates and Costs

A study by Golas et al. (2018) investigated the use of big data to develop a prediction model for all-risk 30-day hospital readmissions of heart failure patients to identify the patients who would benefit from disease management programs and ultimately reduce the rate of readmission and improve overall clinical outcomes (Golas et al., 2018). The study sought out multiple subset models of machine learning and settled on a Deep Unified Networks (DUNs) model, a subset of machine learning. The DUNs model was trained with data sourced from 11,510 patients with 27,334 total admissions, as well as 6,369 readmissions. The DUNs model was found to have an accuracy rate of 76.4% and corresponded with maximum cost savings of 3.403 ± 0.536 (Golas et al., 2018).

Rojas et al. (2018) examined the use of machine learning to derive and validate an intensive care unit (ICU) readmission prediction model and compare it to previously published algorithms, specifically the Stability and Workload Index for Transfer (SWIFT) score and Modified Early Warning Score (MEWS) (Rojas et al., 2018). The study results showed that 11% of the discharges in the study were readmitted to the ICU. (Rojas et al., 2018. This machine-learning model had notably better performance than SWIFT and MEWS. At a specificity of 95%, the model had a sensitivity rate of 28% compared to 15% for SWIFT and 7% for MEWS (Rojas et al., 2018).

Implantable cardiac sensors have shown positive effects in reducing rehospitalization for heart failure, Stehlik et al. (2020) had the objective to determine the accuracy of noninvasive remote monitoring for predicting heart failure rehospitalization (Stehlik et al., 2020). One hundred subjects, 98% male, were enrolled in the study and were studied for up to 3 months using disposable multi-sensor patches where data was uploaded continuously through a smartphone to a cloud analytics system. There were 35 unplanned non-trauma hospitalization events, which included 24 worsening heart failure events. (Stehlik et al., 2020). The platform developed detected precursors of hospitalization for heart failure exacerbation with a 76% – 88% sensitivity and 85% specificity. The reported median time between the initial alert and the readmission was 6.5 days (Stehlik et al., 2020).

A study by Romero-Brufau et al. (2020) focused on a regional Wisconsin hospital between November 2018 and April 2019 to assess patients admitted to general care units for risk of readmission and generated recommendations for interventions intended to decrease readmission risks (Romero-Brufau et al., 2020). The tool assessed 2,460 hospitalizations, among those hospitalizations the tool designated 611 of them as high-risk. The risk assignment sensitivity was 65% and the specificity was 89%. Across the span of the study, the relative readmission rate saw a 25% reduction. (Romero-Brufau et al., 2020).

Big Data for Patient Outcomes

A more recent study used the National COVID Cohort Collaborative (N3C) which contained patient-level data from the original data set of 72 sites across the US, primarily tertiary care centers (Moradi et al., 2023). Moradi et al. (2023) aimed to implement and test three machine learning models to predict the final outcome, death or sufficient improvement to be discharged, of COVID-19 patients. The N3C contained more than 10 million patients and more than 4.9 million SARS-CoV-2 infected persons as of May 4, 2022, the study excluded antibody-only positive results post-December 10, 2022 (Moradi et al., 2023). Following further selection, the finalized set of patients included 145,769 individuals. The Gradient Boosted Decision Tree model presented the best accuracy rate of patient mortality prediction which was 81% (Moradi et al., 2023).

Daghistani, AlGhamdi, Alshammari, and AlHazme (2020) studied the capability of big data and machine learning to predict appointment no-shows, which results in a negative impact on healthcare outcomes (Daghistani et al., 2020). Healthcare outcomes of delay in care have included, but are not limited to, missed medications, mortality of patients with long-term mental and physical health conditions, and overall negative consequences (Henry, 2021; McQueenie et al., 2019; Findling et al., 2020). Appointment no-shows also often lead to a multitude of concerns in addition to patient care, such as increased costs and underutilized resources, prediction techniques would aid in reliable appointment scheduling strategies (Daghistani et al., 2020). Different machine-learning prediction models were evaluated using data from a total of 2,011,813 visits. The Gradient Boosting model was found to have the best performance which resulted in an increase of accuracy to 79% (Daghistani et al., 2020). Figure 3 is displayed to show the accuracy rate of the different models using different validation methods: 70/30 and 80/20 holdout methods and tenfold cross validation. (Daghistani et al., 2020).

Taylor et al. (2015) trained a machine learning predictive analytics model to compare Clinical Decision Rules (CDRs) of emergency care in attempt to predict sepsis mortality while hospitalization. Emergency department visits admitted to the hospital were evaluated if the patient met criteria for sepsis for a year beginning October 2013 (Taylor et al., 2015). The model was constructed using variables from EHRs of four hospitals. 4,676 patients were identified as having met criteria across the span of the study. Multiple models were tested, however, the random forest model showed significantly better results with an 86% confidence model (Taylor et al., 2013).

Big data and machine learning methods have also been developed to predict cardiovascular disease risks based on health record data (Du et al. 2020).Du et al. (2020) used various machine learning methods to analyze EHR data from 42,676 patients with hypertension, in which 20,156 patients had coronary heart disease (CHD) onset (Du et al. 2020). The researchers set out to determine the possibility of getting an accurate risk prediction of CHDan accurate risk prediction for a patient based on their EHR, trained from a mass amount of EHR data. The model used, XGBoost did achieve a high accuracy rate of predicting 3-year CHD onset at 94%. (Du et al. 2020).

DISCUSSION

This study hypothesized that the use of big data has improved quality categories in inpatient facilities led to improved patient outcomes, decreased readmission rates, and reduced costs. The hypothesis was partially confirmed by the positive outcomes shown in the results. The results proved that through the use of big data and machine learning, reducing readmission rates and improving overall patient outcomes were possible. The study was unable to prove cost savings due to a lack of relevant financial statistics and data available.

An interview with an expert in clinical quality was conducted and the corresponding findings have been included in each subcategory of this discussion section.

Reducing Readmission Rates

The findings in this study confirmed the ability of reducing readmission rates through sourcing big data to build machine learning algorithms. The research included in the results showed high rates of accuracy when using prediction models to assess patients' risk of hospitalization and readmission.

<u>Interview:</u> The subject, when asked about the ability of big data to decrease readmission rates, considers the potential for readmission rates to be decreased through human behavior. Suggesting that the model itself does not reduce readmissions, however, the model could direct

clinicians and staff to the patients at risk, but it has required those clinicians to be trained on next steps and to take human action to actually see a reduction in rates Subject Interview, 2023).

Improving Patient Outcomes

Data also showed the effects that big data can have on overall patient outcomes, such as assessment of disease risk and appointment no-shows, which can have a substantial effect on patient's health.

Interview: Patient outcomes could be affected by supporting functions such as staffing appropriately and triaging. The subject suggests that there are many ways to improve patient health with data such as social determinants of health and the use of nurse navigators (Subject Interview, 2023).

Cost Savings

Research included in this study suggested the potential for cost savings through processes being built with machine learning and big data for other tasks and assessments, however, relevant statistics were not available to support this portion of the hypothesis.

Interview: It was found that the interview subject was in agreement with the findings of this research on the topic of cost savings. The subject was overall, in disagreement with the ability to see cost savings in an inpatient setting, but considered that cost savings may be more likely in an outpatient setting (Subject Interview, 2023).

Study Limitations

This study ultimately showed positive results in two portions of the hypothesis, better patient outcomes and reduced readmission rates. There was a limitation to prove the occurrence of cost savings. An additional complication is that, in the research reviewed, and any time when using data analytics there may be potential for the data to not be clean, a limitation mentioned by the interviewee.

Practical Implications

This study provided the examples of use for big data analytics in terms of tools to assist in tasks, clinical or administrative. The mentioned models in the results were previously published studies with a goal to develop a properly functioning machine learning tool, however, little data supports the concept of common use. This said, there has been shown a need for research of inpatient facilities' regular use of machine learning methods, including potential findings of noticeable and correlated cost savings.

CONCLUSION

This research provided valuable insights into the potential for the use of big data in inpatient facilities in the categories of readmission rates and patient outcomes. This study confirmed that with relevant big data and machine learning there are many solutions that can be offered to problems within inpatient facilities. Given the acknowledged study limitations, specifically financial statistics, there is a need for addressing the common use of big data and machine learning tools, the ability for common use, and findings surrounding financial data and cost saving statistics.