4-2024

Navigating the Future of Healthcare: Big Data's Impetus for Inpatient Quality Enhancement

Lakyn Hyre

Brian Cox

Alberto Coustasse

NAVIGATING THE FUTURE OF HEALTHCARE: BIG DATA'S IMPETUS FOR INPATIENT QUALITY ENHANCEMET

Lakyn Hyre
Alumna, Healthcare Administration Program,
Division of Management and Healthcare Administration.
Lewis College of Business
Marshall University
304-746-1968
hare15@marshall.edu\
100 E. Peyton Drive,
South Charleston, WV, 25303

Brian Cox,
DBA student
cox403@marshall.edu
Lewis College of Business
Marshall University
304-746-1968
cox403@marshall.edu

Alberto Coustasse (corresponding Author)
Professor, Healthcare Administration and Health Informatics Program,
Division of Management and Healthcare Administration,
Lewis College of Business
Marshall University
coustassehen@marshall.edu
304-746-1968 (contact author).
100 E. Peyton Drive,
South Charleston, WV, 25303

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

Introduction

In contemporary healthcare, quality improvement remains a critical concern globally, particularly in the United States (US), as evidenced by various scholarly reports (WHO, 2018; ARHQ, 2018). The ramifications of suboptimal quality encompass diverse challenges such as overuse, underuse, and misallocation of services, resulting in avoidable patient complications, increased healthcare costs, and even fatalities (ARQH, 2018). These issues significantly impact financial expenditures, with overtreatment alone accounting for 2%-2.7% of US healthcare spending in 2019, accumulating an estimated $76 billion in additional costs (Health Affairs Research Brief, 2022; Shrank et al., 2019). Concurrently, failures in care delivery and coordination contributed to an estimated $102-166 billion and $27-78 billion in costs, respectively, in the same year (Health Affairs Research Brief, 2022; Shrank et al., 2019).

Among these concerns, hospital readmission rates have become a pivotal indicator of healthcare quality, subsequently influencing healthcare policy, such as the Affordable Care Act's Hospital Readmissions Reduction Program (HRRP) (NEJM Catalyst, 2018b). In 2011, hospital readmissions accounted for 3.3

million cases, imposing $41.3 billion in associated costs (NEJM Catalyst, 2018b). The implementation of HRRP led to a decline in readmission rates from 21.5% in 2007 to 17.8% in 2015 (Zuckerman et al., 2016), reflecting positive strides in quality improvement.

The term "Big Data" has been widely used, but until recently, there was no universally accepted definition for it. According to McKinsey, Big Data refers to datasets too large for typical database software tools to handle effectively. Gartner proposed the "3V" definition: Big Data is characterized by volume, velocity, and variety and requires innovative forms of information processing to gain insights and make decisions. Some definitions also include a fourth dimension called "Veracity," which refers to the data's quality, authenticity, and trustworthiness.

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). Harnessing this data, alongside machine learning, holds promise in enhancing patient outcomes, curbing readmission rates, and potentially reducing costs (Kumar et al., 2018).

Methodology

This study aims to explore the impact of big data utilization on quality improvement in inpatient facilities, specifically assessing its influence on readmission rates, patient outcomes, and potential cost savings. This study utilized mixed methodologies with a literature review complemented by semi-structured interviews to gain perspectives about big data utilization on quality improvement in inpatient facilities. The Marshall University Institutional Review Board (IRB) approved the interview. 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.

Results

Integrating big data analytics and machine learning has significantly transformed healthcare practices, offering predictive models that demonstrate substantial promise in improving patient outcomes and curbing hospital readmission rates.

In the Golas et al. (2018) study, developing a predictive model for 30-day heart failure readmissions utilizing Deep Unified Networks (DUNs) displayed promising accuracy. Trained on data from over 11,000 patients and 27,000 admissions, this model showcased an accuracy rate of 76.4%, highlighting its potential to identify candidates who would benefit from disease management programs, consequently reducing readmission rates and yielding cost savings of up to 3.403 ± 0.536. Rojas et al. (2018) delved into ICU readmission predictions using machine learning, demonstrating superior performance to existing algorithms such as SWIFT and MEWS. Achieving a specificity of 95% and a sensitivity rate of 28%, this model signaled enhanced capabilities in predicting ICU readmissions. Additionally, Stehlik et al. (2020) explored remote monitoring for heart failure using implantable cardiac sensors, which showcased notable sensitivity and specificity in predicting heart failure exacerbation. The platform detected precursors of hospitalization with a sensitivity range of 76% – 88% and a specificity of 85%, significantly aiding in timely interventions and reducing readmission occurrences.

Romero-Brufau et al. (2020) introduced a tool to assess readmission risks for general care unit patients, identifying high-risk hospitalizations. Their tool exhibited a 25% reduction in relative readmission rates, emphasizing its potential to stratify patients and intervene proactively. Furthermore, Moradi et al. (2023) leveraged machine learning on a massive National COVID Cohort Collaborative (N3C) dataset to predict outcomes for COVID-19 patients. With an accuracy rate of 81% in predicting patient mortality, their Gradient Boosted Decision Tree model presented promising results in forecasting patient outcomes.

Daghistani et al. (2020) explored appointment no-show predictions using machine learning techniques. Their Gradient Boosting model achieved an accuracy rate of 79%, assisting in establishing reliable appointment scheduling strategies and addressing concerns related to underutilized resources. Moreover, Taylor et al. (2015) designed a predictive model for sepsis mortality, with the random forest model exhibiting an 86% confidence level in predicting mortality among patients meeting sepsis criteria. Finally, Du et al. (2020) utilized machine learning to predict coronary heart disease onset among hypertensive patients. Their XGBoost model displayed a high % accuracy rate of 94% in predicting 3-year CHD onset based on electronic health record data.

Discussion

The review suggested that big data and machine learning applications have significant potential to achieve the study's objectives. Studies have demonstrated that predictive models accurately identify at-risk patients, enabling proactive interventions to lower readmission rates. The accuracy rates of these models, which range between 70% and 80%, highlight their potential to identify patients requiring additional care post-discharge. These studies illustrated the potential and effectiveness of leveraging big data and machine learning in healthcare. These models contributed to improved patient outcomes and offer promising strategies to mitigate readmission rates across various healthcare domains. However, the study faced limitations in directly confirming associated cost savings. The absence of concrete financial statistics hindered the ability to support the hypothesis regarding cost reductions unequivocally, thus presenting a deviation from the initial aim. While these findings present a promising outlook for integrating big data and machine learning in healthcare, several considerations warrant attention.

Practical Implications

The findings of this research present several practical implications that hold substantial importance for healthcare administrators, clinicians, and policymakers aiming to integrate big data and machine learning in inpatient facilities: Inpatient facilities require a structured strategy for incorporating predictive models into existing workflows. Collaboration among healthcare IT specialists, clinicians, and data scientists is crucial to embedding these models seamlessly into daily clinical practices. Healthcare professionals must receive adequate training to interpret and act upon predictions generated by these models. Targeted resource allocation and proactive patient management can optimize healthcare resources and reduce unnecessary costs. Continuous evaluation and refinement of predictive models is vital, with adherence to regulatory compliance and ensuring patient privacy central to deploying these models. Long-term studies and interdisciplinary collaborations foster innovation and drive the evolution of these models toward more excellent utility and reliability. Developing guidelines and incentives and encouraging research and funding initiatives could provide concrete evidence of the cost-saving potential of these models. In summary, these practical implications underscore the need for meticulous planning, effective integration strategies, continuous evaluation, ethical considerations, and collaborative efforts to harness the transformative potential of big data and machine learning in improving quality and optimizing healthcare outcomes in inpatient facilities.

Limitations

The need for comprehensive financial statistics limits the validation of the proclaimed cost-saving potential of these models. Moreover, challenges about data quality and the seamless integration of these predictive models into routine clinical practices necessitate further exploration.

Future Research Directions

More research is needed to validate the cost-saving benefits of these predictive models. Comprehensive evaluation is necessary for implementing these models in natural healthcare settings. Continued research and practical implementation strategies are crucial for realizing the potential of these innovative healthcare approaches.

Conclusion

In conclusion, while affirming the transformative potential of big data and machine learning in enhancing inpatient facility quality and patient outcomes, this study underscores the need for further research to validate cost-saving benefits and address implementation challenges.

References

Abouelmehdi, K., Beni-Hessane, A. & Khaloufi, H. (2018, January 9). *Big healthcare data: preserving security and privacy. Journal of Big Data* 5 (1). Retrieved on October 7, 2023 from https://doi.org/10.1186/s40537-017-0110-7.

Agency for Healthcare Research and Quality (AHRQ). (2018, June) *The challenge and potential for assuring quality health care for the 21st Century*. Retrieved August 29, 2023 from https://www.ahrq.gov/patient-safety/quality-measures/21st-century/index.html#:~:text=There%20are%20several%20areas%20where,higher%20costs%2C%20and%20premature%20deaths.

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Affairs*, *33*(7), 1123-31. DOI:10.1377/hlthaff.2014.0041 .

Centers for Medicare and Medicaid Services (CMS). (2021, December 1). *Quality Measurement and Quality Improvement*. Retrieved August 28, 2023 from https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/mms/quality-measure-and-quality-improvement-

Daghistani, T., AlGhamdi, H., Alshammari, R., Alhazme, RH. (2020, December 9).Predictors of outpatients' no-show: big data analytics using Apache spark. *Journal of Big Data* 7, 108. https://doi.org/10.1186/s40537-020-00384-9

Du, Z., Yang, Y., Zheng, J., Li, Q., Lin, D., Li, Y., Fan, J., Cheng, W., Chen, X.-H., & Cai, Y. (2020, July 6). Accurate prediction of coronary heart disease for patients with hypertension from electronic health records with Big Data and machine-learning methods: Model Development and Performance Evaluation. *JMIR medical informatics.8*(7), e17257. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7381262/

Findling, M., Blendon, R., Benson, J. (2020, December 14) Delayed Care with Harmful Consequences-Reported Experiences From National Surveys During Coronavirus Disease 2019. *JAMA Health Forum. 1*(12), e201463. doi:10.1001/jamahealthforum.2020.1463

Fischer, C., Lingsma, H. F., Marang-van de Mheen, P. J., Kringos, D. S., Klazinga, N. S., & Steyerberg, E. W. (2014). Is the readmission rate a valid quality indicator? A review of the evidence. *PloS one*, *9*(11), e112282. https://doi.org/10.1371/journal.pone.0112282

Golas, S.B., Shibahara, T., Agboola, S., Otaki, H., Sato, J., Nakae, T.,...Jethwani, K. (2018, June 22). A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC Medical Informatics and Decision Making. 18*(1),44 https://doi.org/10.1186/s12911-018-0620-z

Health Affairs Research Brief. (2022, June 9) *The Role Of Clinical Waste In Excess US Health Spending.* DOI: 10.1377/hpb20220506.432025

Henry, T. (2021, February 15) Why 41% of patients have skipped care during COVID-19 pandemic. *American Medical Association.* https://www.ama-assn.org/delivering-care/public-health/why-41-patients-have-skipped-care-during-covid-19-pandemic

Kumar S, Singh M. (2019, March) Big Data Analytics for Healthcare Industry: Impact, Applications, and Tools. *Big Data Mining and Analytics*, 2(1): 48-57. https://doi.org/10.26599/BDMA.2018.9020031

McQueenie, R., Ellis, D. A., McConnachie, A., Wilson, P., & Williamson, A. E. (2019). Morbidity, mortality and missed appointments in healthcare: a national retrospective data linkage study. *BMC Medicine, 17*(1), doi:10.1186/s12916-018-1234-0

Moradi, H., Bunnell, H. T., Price, B. S., Khodaverdi, M., Vest, M. T., Porterfield, J. Z., … Hodder, S. L. (2023, March 9). Assessing the effects of therapeutic combinations on SARS-COV-2 infected patient outcomes: A big data approach. *PLOS ONE*, *18*(3). doi:10.1371/journal.pone.0282587

NEJM Catalyst. (2018a, January 1). *Healthcare Big Data and the Promise of Value Based Care.* Retrieved on October 7, 2023 from https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0290

NEJM Catalyst. (2018b, April). *Hospital Readmissions Reduction Program (HRRP).* Retrieved on October 7, 2023, from https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0194

Rojas, J. C., Carey, K. A., Edelson, D. P., Venable, L. R., Howell, M. D., & Churpek, M. M. (2018, July). *Predicting Intensive Care Unit Readmission with Machine Learning Using Electronic Health Record Data. Annals of the American Thoracic Society, 15*(7), 846-853. https://doi.org/10.1513/AnnalsATS.201710-787OC

Romero-Brufau, S., Wyatt, K. D., Boyum, P., Mickelson, M., Moore, M., & Cognetta-Rieke, C. (2020, September 2). Implementation of Artificial Intelligence-Based Clinical Decision Support to Reduce Hospital Readmissions at a Regional Hospital. *Applied clinical informatics*, *11*(4), 570–577. https://doi.org/10.1055/s-0040-1715827

Shrank, W. H., Rogstad, T. L., & Parekh, N. (2019, October 7). *Waste in the US Health Care System*. JAMA. DOI:10.1001/jama.2019.13978

Stehlik, J., Schmalfuss, C., Bozkurt, B., Nativi-Nicolau, J., Wohlfahrt, P., Wegerich, S., . . . Pham, M. (2020, February 25). Continuous Wearable Monitoring Analytics Predict Heart Failure Hospitalization. *Circulation: Heart Failure, 13*(3), e006513. DOI:10.1161/CIRCHEARTFAILURE.119.006513

Subject Interview. Conducted by Lakyn Hare. (2023, November 16).

Taylor, AR., Pare, JR.,Venakatesh, AK., Mowafi, H., Melnick, ER., Fleischman, W., Hall, K. (2015, December 17) Prediction of in-hospital mortality in emergency department patients with sepsis: A local big data-driven, machine learning approach. *Academic Emergency Medicine*, *23*(3), 269-278. **https://doi.org/10.1111/acem.12876**

Tulane University. (2021, July 7). *Big data in health care and patient outcomes*. Retrieved August 28, 2023 from https://publichealth.tulane.edu/blog/big-data-in-healthcare/

World Health Organization (WHO). (2018, July 5) *Low Quality Healthcare is increasing the burden of illness and health costs globally*. Retrieved August 28, 2023 from https://www.who.int/news/item/05-07-2018-low-quality-healthcare-is-increasing-the-burden-of-illness-and-health-costs-globally

Yao, W., Chao-Hsien, C., Li, Z. (2010) The Use of RFID in Healthcare: Benefits and Barriers. Proceedings of the 2010 IEEE International Conference on RFID-Technology and Applications (RFID-TA), 128–34.

Zuckerman, R. B., Sheingold, S. H., Orav, E. J., Ruhter, J., & Epstein, A. M. (2016,). Readmissions, Observation, and the Hospital Readmissions Reduction Program. *The New England journal of medicine*, *374*(16), 1543–1551. https://doi.org/10.1056/NEJMsa1513024

Grande, D., Machado, J., Petzold, B., & Roth, M. (2020, December 16). How to reduce data costs while increasing speed and quality. *McKinsey & Company*. 1 retrieved on December 26, 2023 from Reducing data costs without sacrificing growth | McKinsey

Manyika J, Chui M, Brown B, et al. Big data: The next frontier for innovation, competition, and productivity. 2011. www.mckinsey.com/mgi (12 September 2019,

Beyer M, Laney D The Importance of 'Big Data': A Definition. Gartner report. 2012, 1–9.

Figures and tables

Figure 1: Conceptual Framework, adapted from (Yao, Chu, Li., 2010)

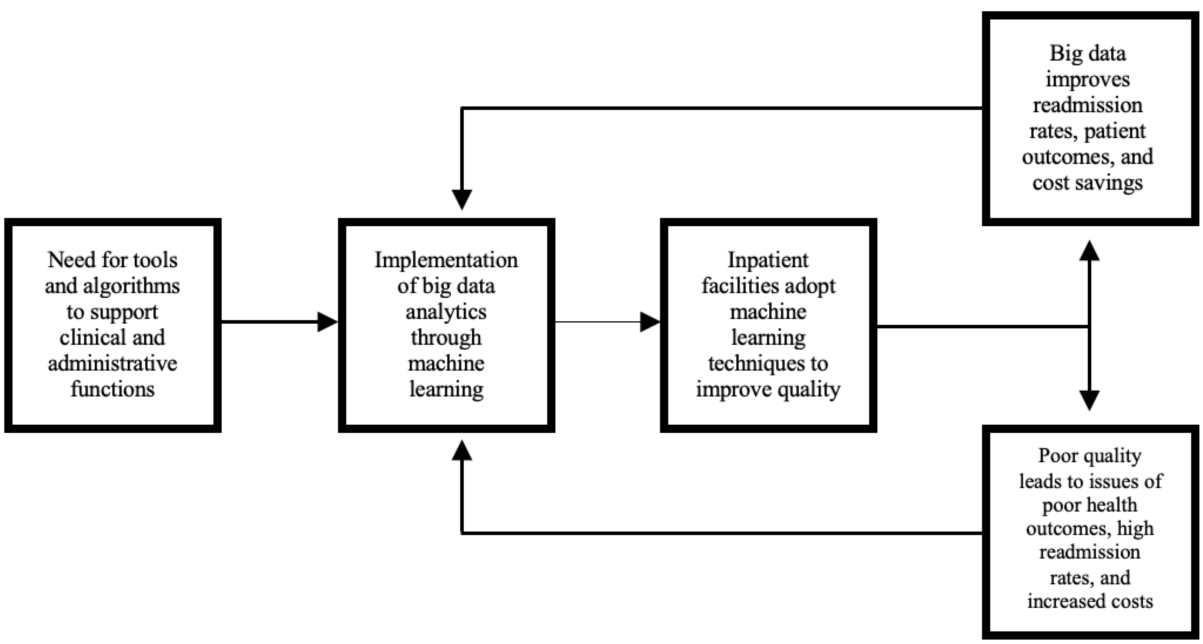

Table 1: Accuracy of different models on predicting outpatient no-show using different alidation methods. Data retrieved from (Daghistani et al., 2020).

Model	Accuracy (70/30 holdout method)	Accuracy (80/20 holdout method)	Accuracy (tenfold cross validation)
Random Forest	0.76	0.75	0.76
Gradient Boosting	0.79	0.79	0.79
Logistic Regression	0.75	0.75	0.75
SVM	0.73	0.54	0.73
Multilayer Perceptron	0.77	0.75	0.77