

Data-Driven Insights in Healthcare

Victor R. Prybutok ^{1,*}  and Gayle L. Prybutok ² 

¹ Department of Information Technology and Decision Sciences, G. Brint Ryan College of Business, University of North Texas, Denton, TX 76203, USA

² Department of Rehabilitation and Health Services, College of Health and Public Service, University of North Texas, Denton, TX 76203, USA

* Correspondence: victor.prybutok@unt.edu; Tel.: +01-940-565-4767

1. Introduction

We are pleased to present this Special Issue, which is a curated collection of research that showcases the transformative power of data-driven approaches in healthcare. The healthcare sector generates vast amounts of observational data daily, yet systematic exploration of these datasets to uncover meaningful patterns remains underutilized. The rapid advancement of digital health technologies, including electronic health records, medical imaging systems, wearable devices, and genomic sequencing platforms, has led to an exponential growth in healthcare data availability [1]. In addition, data from routine clinical practices offer unique opportunities to complement evidence from randomized controlled trials, particularly for understanding treatment effectiveness in diverse patient populations and real-world clinical settings [2,3]. However, ensuring the quality and appropriate use of these observational datasets remains a persistent challenge that requires systematic attention [4,5]. The collection in this Special Issue demonstrates that while theory-driven research remains essential, data exploration can generate hypotheses, reveal unexpected associations, and provide evidence that directly informs practice and future research directions.

The ten contributions assembled in this Special Issue span diverse healthcare contexts and analytical methodologies, unified by a commitment to extracting actionable knowledge from empirical observations. These publications employ innovative techniques, including artificial intelligence, machine learning, natural language processing, advanced statistical modeling, operations research, and exploratory data analytics, to address critical challenges in healthcare delivery, quality improvement, and patient outcomes [6]. The integration of these data-driven methodologies supports the need for future research and application regarding how healthcare systems can leverage observational evidence to inform clinical decision-making and policy development [7].

2. Artificial Intelligence and Machine Learning for Clinical Prediction

The Special Issue opens with three contributions that demonstrate the power of artificial intelligence and advanced analytical methods for clinical prediction and decision support. Halwani and Halwani (contribution 1) present “Prediction of COVID-19 Hospitalization and Mortality Using Artificial Intelligence,” employing decision trees, support vector machines, and random forest algorithms to predict hospital mortality among COVID-19 patients. Their analysis of data from King Abdulaziz University Hospital in Saudi Arabia achieved predictive accuracy rates of 76–82%, with hospital stay duration, D-Dimers, alkaline phosphatase, bilirubin, lactate dehydrogenase, C-reactive protein, and ferritin identified as significant mortality predictors. This work illustrates how AI tools



Received: 14 October 2025

Revised: 17 October 2025

Accepted: 20 October 2025

Published: 22 October 2025

Citation: Prybutok, V.R.; Prybutok, G.L. Data-Driven Insights in Healthcare. *Healthcare* **2025**, *13*, 2658. <https://doi.org/10.3390/healthcare13212658>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

can enhance early identification of high-risk patients and support clinical decision-making during pandemic situations.

Alasmari (contribution 2) provides a comprehensive scoping review titled “A Scoping Review of Arabic Natural Language Processing for Mental Health,” examining NLP techniques applied to mental health detection in Arabic-speaking populations. Following the PRISMA-ScR framework, this review identifies the effectiveness of various approaches, with transformer-based models such as AraBERT and MARBERT achieving superior performance with accuracy rates up to 99.3% and 98.3%, respectively. The review highlights how NLP can analyze social media data to detect depression and suicidality, demonstrating the potential of these techniques for population-level mental health surveillance in linguistically diverse contexts.

Chang, Ryu, Choi, Kwon, and Kim (contribution 3) present “A Comparative Study of Hospitalization Mortality Rates between General and Emergency Hospitalized Patients Using Survival Analysis,” employing Kaplan–Meier survival estimation and Cox proportional hazards models to analyze four years of data from the Korean National Health Insurance Services. Their analysis reveals distinct determinants of mortality risk between general inpatients and emergency admissions, with geographic factors and institutional characteristics such as physician and nurse staffing ratios, bed capacity, and emergency bed availability showing differential effects. This work demonstrates how survival analysis techniques can accommodate censored medical data characteristics often overlooked by conventional regression approaches.

3. Data-Driven Quality Improvement and Healthcare Standards

Two articles explore how data-driven methodologies can enhance healthcare quality standards and establish evidence-based benchmarks. Richardson, Penumaka, Smoot, Panaganti, Chinta, Guduri, Tiyyagura, Martin, Korvink, and Gunn (contribution 4) present “A Data-Driven Approach to Defining Risk-Adjusted Coding Specificity Metrics for a Large U.S. Dementia Patient Cohort,” analyzing 487,775 hospitalization records to develop risk-adjusted metrics for assessing medical coding specificity. Using logistic regression models incorporating patient and facility characteristics, combined with Poisson binomial modeling, they created benchmarks enabling healthcare facilities to assess coding practices against industry standards. With an AUC of 0.727 for principal dementia diagnoses, their approach demonstrates how data-driven methods can identify facilities that over- or under-specify diagnoses, ultimately contributing to improved patient care quality and healthcare system reliability.

Velev, Velazquez-Sosa, Lebien, Janwa, and Roche-Lima (contribution 5) provide “Modeling Multivariate Distributions of Lipid Panel Biomarkers for Reference Interval Estimation and Comorbidity Analysis,” employing Gaussian Mixture Models to derive reference intervals directly from large-scale, real-world laboratory data from Puerto Rico. Their methodology enables separation of healthy and pathological subpopulations without relying on diagnostic codes, producing sex- and age-stratified reference intervals for total cholesterol, LDL, HDL, and triglycerides. By examining selective mortality patterns and constructing comorbidity implication networks, they explain counterintuitive age trends in lipid values and characterize interdependencies between conditions, demonstrating how population-specific reference intervals can be derived without recruiting healthy cohorts.

4. Healthcare Operations Research and Resource Optimization

Two contributions employ operations research methodologies to optimize healthcare resource allocation and workforce management. Mystakidis, Koukaras, Koukaras, Kaparis, Stavrinides, and Tjortjis (contribution 6) present “Optimizing Nurse Rostering:

A Case Study Using Integer Programming to Enhance Operational Efficiency and Care Quality,” developing a comprehensive integer programming model for nurse scheduling in oncology departments. Their model integrates constraints including legal work hours, staff qualifications, and personal preferences to generate equitable and efficient schedules. Implementation in a clinical setting revealed significant improvements in scheduling efficiency, staff satisfaction, workload distribution, and compliance with work-hour regulations, demonstrating how operations research techniques can enhance both operational excellence and staff well-being in acute care settings.

Clapper, ten Hove, Bekker, and Moeke (contribution 7) provide “Team Size and Composition in Home Healthcare: Quantitative Insights and Six Model-Based Principles,” developing six model-based principles to guide managerial decisions regarding home healthcare team structure. Through extensive data analysis and mathematical modeling based on real-life scenarios, they demonstrate that efficiency improves with team size but with diminishing returns, while team manageability becomes increasingly complex as size grows. Their work provides estimates for travel time based on team size and territory, establishes upper bounds for full-time contract fractions to avoid split shifts, and concludes that ideally sized teams should serve at least several hundred care hours weekly. This research exemplifies how quantitative modeling can inform practical workforce planning decisions.

5. Technology-Enabled Healthcare and Behavioral Insights

Two articles examine how technology shapes health-seeking behaviors and social interactions in healthcare contexts. Boyce, Harun, G. Prybutok, and V. Prybutok (contribution 8) present “The Role of Technology in Online Health Communities: A Study of Information-Seeking Behavior”, employing partial least squares structural equation modeling with multi-group and importance-performance map analysis to examine technology’s role in online health communities. Their cross-sectional survey identifies ease of site navigation and interaction with other members as the most beneficial technology-related factors influencing information-seeking processes. The findings provide actionable insights for developing and managing online health communities and for healthcare professionals seeking to disseminate relevant information to individuals with chronic illnesses such as COPD.

Chen, Hsu, and Rahman (contribution 9) provide “From Mandate to Choice: How Voluntary Mask Wearing Shapes Interpersonal Distance Among University Students After COVID-19,” examining the association between voluntary protective behaviors and social interactions in post-mandate Taiwan. Through an online interpersonal distance simulation with 100 university students, they employed four-way ANOVA to reveal that mask-wearing individuals maintain significantly greater interpersonal distances, suggesting heightened risk perception, while masked targets elicit smaller distances, possibly due to safety signaling. Gender differences emerged in both protective behavior adoption (72% of females versus 44% of males) and spatial preferences, offering insights into how voluntary behavioral adaptations continue shaping social norms after mandate removal.

6. Research Infrastructure and Data Governance

The Special Issue concludes with a perspective on research infrastructure challenges. Landi, D’Ambrosio, Faggion, Rocchi, Paganin, Lain, Ceci, and Giannuzzi on behalf of the EPIICAL Consortium (contribution 10) present “Sharing Data and Transferring Samples Within Pediatric Clinical Studies: How to Overcome Challenges and Make Them a Science Opportunity.” This perspective examines the EPIICAL project’s establishment of a dedicated Working Group to navigate ethical and regulatory complexities of international pediatric clinical studies involving HIV-infected children. The consortium developed well-

structured informed consent and assent templates, data sharing agreements, and material transfer agreements to regulate sample transfers among partners and sites across European and non-European boundaries. This contribution highlights how structured governance frameworks and expert support can transform regulatory challenges into opportunities for advancing pediatric clinical research.

7. Conclusions

The contributions assembled in this Special Issue collectively demonstrate that data-driven discovery in healthcare extends far beyond descriptive analysis. These works show how systematic exploration of observational data using diverse methodologies, from artificial intelligence and machine learning to operations research and behavioral modeling, can generate actionable insights that improve patient care, enhance operational efficiency, establish evidence-based standards, and inform policy decisions.

As healthcare systems continue generating unprecedented volumes of data through digital transformation initiatives [1,6], the approaches showcased in this Special Issue provide both inspiration and practical guidance for researchers, practitioners, and policy-makers seeking to extract maximum value from their data assets. The successful translation of data-driven insights into improved patient outcomes requires not only sophisticated analytical methods but also robust data governance frameworks, quality assurance mechanisms, and ethical oversight [2,4]. We hope this collection serves as a catalyst for continued innovation in data-driven healthcare research and practice.

Author Contributions: V.R.P. and G.L.P. contributed equally to the conceptualization, curation, and writing of this editorial. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

List of Contributions

1. Halwani, M.A.; Halwani, M.A. Prediction of COVID-19 Hospitalization and Mortality Using Artificial Intelligence. *Healthcare* **2024**, *12*, 1694. <https://doi.org/10.3390/healthcare12171694>.
2. Alasmari, A. A Scoping Review of Arabic Natural Language Processing for Mental Health. *Healthcare* **2025**, *13*, 963. <https://doi.org/10.3390/healthcare13090963>.
3. Chang, H.; Ryu, S.; Choi, I.; Kwon, A.E.; Kim, J. A Comparative Study of Hospitalization Mortality Rates between General and Emergency Hospitalized Patients Using Survival Analysis. *Healthcare* **2024**, *12*, 1982. <https://doi.org/10.3390/healthcare12191982>.
4. Richardson, K.; Penumaka, S.; Smoot, J.; Panaganti, M.R.; Chinta, I.R.; Guduri, D.P.; Tiyyagura, S.R.; Martin, J.; Korvink, M.; Gunn, L.H. A Data-Driven Approach to Defining Risk-Adjusted Coding Specificity Metrics for a Large U.S. Dementia Patient Cohort. *Healthcare* **2024**, *12*, 983.
5. Velez, J.; Velázquez-Sosa, L.; Lebien, J.; Janwa, H.; Roche-Lima, A. Modeling Multivariate Distributions of Lipid Panel Biomarkers for Reference Interval Estimation and Comorbidity Analysis. *Healthcare* **2025**, *13*, 2499. <https://doi.org/10.3390/healthcare13192499>.
6. Mystakidis, A.; Koukaras, C.; Koukaras, P.; Kaparis, K.; Stavrinides, S.G.; Tjortjis, C. Optimizing Nurse Rostering: A Case Study Using Integer Programming to Enhance Operational Efficiency and Care Quality. *Healthcare* **2024**, *12*, 2545. <https://doi.org/10.3390/healthcare12242545>.
7. Clapper, Y.; ten Hove, W.; Bekker, R.; Moeke, D. Team Size and Composition in Home Healthcare: Quantitative Insights and Six Model-Based Principles. *Healthcare* **2023**, *11*, 2935. <https://doi.org/10.3390/healthcare11222935>.
8. Boyce, L.; Harun, A.; Prybutok, G.; Prybutok, V.R. The Role of Technology in Online Health Communities: A Study of Information-Seeking Behavior. *Healthcare* **2024**, *12*, 336. <https://doi.org/10.3390/healthcare12030336>.
9. Chen, Y.-L.; Hsu, C.-W.; Rahman, A. From Mandate to Choice: How Voluntary Mask Wearing Shapes Interpersonal Distance Among University Students After COVID-19. *Healthcare* **2025**, *13*, 1956. <https://doi.org/10.3390/healthcare13161956>.

10. Landi, A.; D'Ambrosio, F.; Faggion, S.; Rocchi, F.; Paganin, C.; Lain, M.G.; Ceci, A.; Giannuzzi, V., on behalf of the EPIICAL Consortium. Sharing Data and Transferring Samples Within Pediatric Clinical Studies: How to Overcome Challenges and Make Them a Science Opportunity. *Healthcare* **2024**, *12*, 2473. <https://doi.org/10.3390/healthcare12232473>.

References

1. Rahman, M.A.; Moayedikia, A.; Wiil, U.K. Editorial: Data-Driven Technologies for Future Healthcare Systems. *Front. Med. Technol.* **2023**, *5*, 1183687. [[CrossRef](#)] [[PubMed](#)]
2. Lighterness, A.; Adcock, M.; Scanlon, L.A.; Price, G. Data Quality-Driven Improvement in Health Care: Systematic Literature Review. *J. Med. Internet Res.* **2024**, *26*, e57615. [[CrossRef](#)] [[PubMed](#)]
3. Barbieri, D.; Chudy-Onwugaje, K.; Langel, J.; Peluso, M.J.; Torres, L.; Kelly, J.D.; Deter, H. Real-World Data and Real-World Evidence in Healthcare in the United States and European Union. *Bioengineering* **2024**, *11*, 784. [[CrossRef](#)]
4. Alam, M.A.; Sajib, M.R.U.Z.; Rahman, F.; Ether, S.; Hanson, M.; Sayeed, A.; Akter, E.; Nusrat, N.; Islam, T.T.; Raza, S.; et al. Implications of Big Data Analytics, AI, Machine Learning, and Deep Learning in the Health Care System of Bangladesh: Scoping Review. *J. Med. Internet Res.* **2024**, *26*, e54710. [[CrossRef](#)] [[PubMed](#)]
5. Ricotta, E.E.; Rid, A.; Cohen, I.G.; Evans, N.G. Observational Studies Must Be Reformed Before the Next Pandemic. *Nat. Med.* **2023**, *29*, 1903–1905. [[CrossRef](#)] [[PubMed](#)]
6. Elragal, A.; Elragal, H.; Habibipour, A. Healthcare Analytics—A Literature Review and Proposed Research Agenda. *Front. Big Data* **2023**, *6*, 1277976. [[CrossRef](#)] [[PubMed](#)]
7. Gershon, A.S.; Lindenauer, P.K.; Wilson, K.C.; Rose, L.; Walkey, A.J.; Sadatsafavi, M.; Anstrom, K.J.; Au, D.H.; Bender, B.G.; Brookhart, M.A.; et al. Informing Healthcare Decisions with Observational Research Assessing Causal Effect: An Official American Thoracic Society Research Statement. *Am. J. Respir. Crit. Care Med.* **2021**, *203*, 14–23. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.