EDITORIAL

Referência
REVISTA DE ENFERMAGEM | JOURNAL OF NURSING

Unveiling the Hidden Potential of Big Data: Empowering Nurses through Data Science

Revelar o potencial oculto dos megadados: Capacitar os enfermeiros através da ciência de dados

Vinciya Pandian [1]
https://orcid.org/0000-0002-2260-1080

Michael J. Brenner [2, 3]
https://orcid.org/0000-0003-4926-0957

[1] Johns Hopkins School of Nursing, Baltimore, United States

[2] University of Michigan School of Medicine, Department of Otolaryngology – Head & Neck Surgery, Michigan, United States

[3] Global Tracheostomy Collaborative, Illinois, Chicago, United States

Corresponding author
Vinciya Pandian
E- mail: vpandia1@jhu.edu

Introduction

Over the past millennium, the landscape of healthcare has undergone a profound metamorphosis from reliance on intuition and traditional remedies to evidence-based practice that leverages data science. The 19th and 20th centuries witnessed significant strides in evidence-based medicine, with the advent of scientific research and the systematic collection of data, and the integration of technology in the digital age that has ushered in a new era data management, with electronic health records (EHRs) allowing for standardization and aggregation of patient-level data at scale. The synergy of big data and data science has emerged and is revolutionizing healthcare, and this Renaissance affords unprecedented insights into patient care, treatment outcomes, and population health. Nursing is poised to leverage these opportunities, utilizing sophisticated technologies and analytics to deliver personalized and effective patient care. The intersection of nursing, big data, and data science marks a paradigm shift towards a more data-driven and interconnected healthcare system capable of enhancing the quality of patient outcomes. In this editorial, we trace the history, present state, and future of data science, shining a light on how the nursing profession can lead in utilizing digital innovation to catalyze progress in education, population health, drug discovery, and health equity; we also consider the ethical dimensions of privacy, data integrity, and bias. As trailblazers in evidence-based practice, interdisciplinary collaboration, and educational innovation, nurses are destined to shape the future of digital healthcare.

Historical perspectives of clinical documentation

The roots of clinical documentation can be traced back to ancient civilizations. In Ancient Mesopotamia, clay tablets inscribed with cuneiform writing provide evidence of some of the earliest known medical records (Lorkowski & Pokorski, 2022). In Egypt, medical information was transcribed onto scrolls made of papyrus, while in Greece and Rome, parchment was used for documenting medical practices (Lorkowski & Pokorski, 2022). The transition from these ancient practices to more sophisticated methods occurred gradually over the centuries. With the advent of the printing press, medical knowledge became more widely disseminated, allowing for standardized documentation. However, the rise of Electronic Health Record (EHR) systems transformed the structure of not only healthcare documentation but also all aspects of communication and healthcare delivery; it has empowered healthcare professionals to store, manage, and retrieve patient information and provided the digital architecture and repository necessary for data science (Dicuonzo et al., 2022). Big data is characterized by extensive datasets that surpass the capabilities of conventional database software, and harnessing this data can transform healthcare (Batko & Ślęzak, 2022). Computational analysis can uncover patterns, trends, and associations in patient care, detecting signals within the data that are only discernable with population data. This evolution signifies a broader and more comprehensive

approach to extracting meaningful insights from clinical documentation. Data science incorporates advanced analytical techniques, machine learning algorithms, and statistical models to derive actionable information from complex datasets. The shift in terminology from "big data" to "data science" reflects a recognition that it is not only volume of data that matters but also the analytical methodologies and tools applied to it. Data science encompasses predictive analytics, personalized medicine, and population health management (Subrahmanya et al., 2022), and data science is shaping the future of healthcare by unlocking insights that can improve patient outcomes, streamline processes, and enhance healthcare delivery.

Characteristics of big data

Big Data has several dimensions that can be summarized by the 5 v's of *volume*, *velocity*, *variety*, *veracity*, and *value*, and each of the dimensions is applicable to nursing practice. *Volume* in data science entails extensive datasets from patient records to medical imaging files, often measured in terabytes (Batko & Ślęzak, 2022). *Velocity* reflects rate of accrual of patient data, often necessitating real-time processing to inform clinical decisions (Batko & Ślęzak, 2022). *Variety* encompasses the format and source of data, including structured patient records, semi-structured data from wearable devices, and unstructured data from clinical narratives (Batko & Ślęzak, 2022). *Veracity* relates to the accuracy and reliability of patient information, which is critical to uphold the quality of care (Batko & Ślęzak, 2022). Finally, *value* underscores the primary goal of extracting meaningful insights from nursing data, facilitating continuous improvements in patient care protocols and healthcare delivery
(Batko & Ślęzak, 2022).

Importance of data Science in healthcare

Data science is relevant in diverse areas of nursing from clinician decision support to predictive algorithms to anticipate and manage disease outbreaks. Data science can accelerate drug discovery, identify operational inefficiencies, and support preventive measures to benefit population health (Yogesh & Karthikeyan, 2022). Despite the promise of this technology, challenges persist in leveraging big data for research and reporting quality outcomes. Nurses must consistently document and collect data to support research, fostering collaboration and knowledge translation. Florence Nightingale's pioneering use of statistics during the Crimean War underscores the transformative power of data (Bradshaw, 2020). She demonstrated how evidence, not perception or tradition, should guide decision-making. Nurses today have the opportunity to harness data science for improved health outcomes (Bradshaw, 2020). Data science, facilitated by handheld devices and wearable technologies, is transforming healthcare (Uddin & Syed-Abdul, 2020). Individualized models based on personal health trajectories enable personalized care. Amid the rapid influx of data from data warehouses, wearables, and other sources, nurses play a key role in meaningful analyses at both individual and population levels.

Data analytics in nursing

According to the National Institute of Standards and Technology, analytics involves the systematic processing and manipulation of data to unveil patterns, relationships, historical trends, and predictions of future behaviors and events (National Institute of Standards and Technology, 2015). Data analytics is thus far broader than statistical analyses; the field encompasses information preparation, pattern analysis, knowledge synthesis, and value creation (National Institute of Standards and Technology, 2015). The analytics process commences with data collection, extraction, processing, transformation, and proceeds to analysis, model and algorithm creation, culminating in interpretation and reporting that can inform decision-making (National Institute of Standards and Technology, 2015). Analytics encompasses the methods, implementation, and outcomes necessary to synthesize information and drive decision-making.

Types of analytics

Analytics can be descriptive, diagnostic, predictive, or prescriptive; *Descriptive Analytics* include use of business intelligence or data mining to explain past occurrences, providing insights into historical events without delving into underlying causes (Sarker, 2021); *Diagnostic Analytics* seeks to uncover the reasons behind events, comparing historical data against other sources and providing in-depth insights into specific problems (Sarker, 2021); *Predictive Analytics* utilize statistical models and forecasts to predict future trends, often based on machine learning or deep learning algorithms (Sarker, 2021); *Prescriptive Analytics* employs optimization and simulation to recommend actions, aiming to eliminate future problems or capitalize on promising trends. It requires sophisticated tools, technologies, and careful consideration of historical and external data (Sarker, 2021).

Existing sources for data science in nursing

In the realm of data science in nursing, an array of existing sources proves indispensable for nurses seeking valuable insights. MarketScan, a longitudinal database, offers individual-level healthcare claims, facilitating in-depth studies on surgical care and disease management (Hirsch & Pitak-Arnnop, 2023; Premo et al., 2023; Wadhwa et al., 2023; Xie et al., 2023). The Healthcare Cost and Utilization Project (HCUP) provides state and national databases, of-

fering comprehensive information on inpatient hospital discharges, emergency department visits, and ambulatory surgery center encounters (Ma et al., 2023; Matabele et al., 2023). Nationwide databases such as the Kids' Inpatient Database, National Inpatient Sample, and National Readmission Databases contribute family-based inpatient data, aiding in national disease burden estimates (Metcalfe et al., 2019). Specialized registries, like Reg-ent clinical data registry maintained by the American Academy of Otolaryngology-Head & Neck Surgery, offer clinical data specific to specialties, enabling outcome assessments (Saraswathula et al., 2023; Schmalbach et al., 2021). Furthermore, initiatives like the Global Tracheostomy Collaborative Registry collect data from over 50 hospitals globally, utilizing data science to enhance the understanding and treatment of tracheostomy patients (Brenner et al., 2020). These diverse sources collectively empower nurses in their data science endeavors, providing a rich foundation for analysis and innovation in healthcare.

Challenges of data Science in healthcare

Ethical considerations are a cornerstone of leveraging big data and data science for advancements in healthcare. Specific considerations relate to data privacy, security, and equity. Stakeholders -ranging from patients to clinicians - must adopt an open-minded perspective that includes a willingness to critically examine and learn from data science. Ensuring accuracy and consistency of data are pivotal, necessitating continuous optimization to avoid erroneous insights and decisions (Osama et al., 2023). Additionally, working with diverse data sets requires navigating data silos, interoperability issues, and mitigating bias (Osama et al., 2023). Integrating disparate data storage systems requires mapping, transformation, consolidation, and ensuring compatibility for meaningful analysis and decision--making (Osama et al., 2023). Such efforts are critical for representative analysis that can help mitigate bias and perpetuating health inequity.

Technological advances and emerging trends in data science

Data science empowers nurses to deliver personalized care and advocate for progress in healthcare practices. Predictive analytics allow for proactive identification of impending health concerns, and the expansion of telehealth and remote monitoring can improve monitoring chronic conditions. The rise of wearable technology adds another dimension, empowering nurse to use data from devices and sensors to track vital signs, activity levels, and treatment plan adherence (Uddin & Syed-Abdul, 2020). Genomic medicine is another frontier for data science, allowing nurses to develop and implement personalized treatment plans (Hassan et al., 2022). Data science is also instrumental in population health management, allowing nurses to identify at-risk populations and to design targeted interventions to improve community health. Natural Language Processing can extract data from unstructured clinical notes and narratives, facilitating deeper understanding (Aramaki et al., 2022).

As the field advances, nurses must recognize the promise and perils of ever-expanding datasets and analytic capabilities. Cybersecurity is critical for ensuring the privacy of patient data, and the there is need for constant vigilance to safeguard sensitive information in the healthcare sector (Cremer et al., 2022). The advent of precision medicine, genomic testing, and the integration of Artificial Intelligence (AI) into data analytics are transforming nursing care (Johnson et al., 2021). Clinical trials are increasingly conducted virtually. Smartphones allow for continuous trial access, and the increasing computing capabilities of devices have opened new possibilities. Machine learning is a major catalyst in the evolution of data science, laying the foundation for increasing integration of AI into healthcare. Nurses will be key players in embracing change and adapting to new frameworks that are necessary to navigate the ever-changing landscape of healthcare.

Data and the future of nursing

By embracing the integration of big data into healthcare, nurses can lead in groundbreaking discoveries and the development of data-driven tools that revolutionize patient care. Data science can provide the foundation for immersive technologies necessary for accelerating learning and enhancing performance. Data science empowers future healthcare professionals to navigate the complexities of a data-rich healthcare landscape with confidence and competence. Nurses are instrumental in policy discussions and can advocate for ethical data use, patient privacy, and responsible use of big data. Interdisciplinary collaboration with professionals from diverse fields fosters innovation and leads to solutions that transcend traditional boundaries. By embracing lifelong learning and staying abreast of advances in data science, nurses will be prepared to meet the challenges of the future.

Conclusion

The transformative role of big data and data science in healthcare has positioned nurse as leader in shaping the future of healthcare. As nurses leverage technological advances in the context of education, interdisciplinary collaboration, and ethical data usage, they will lead in research, innovation, and advocacy. Doing so requires developing proficiency in accessing data, interpreting data analytics, and embracing emerging trends. Nurses are positioned to revolutionize patient care through the judicious application of big data and data science.

References

Aramaki, E., Wakamiya, S., Yada, S., & Nakamura, Y. (2022). Natural Language Processing: from Bedside to Everywhere. *Yearb Med Inform*, *31*(1), 243-253. https://doi.org/10.1055/s-0042-1742510

Batko, K., & Ślęzak, A. (2022). The use of Big Data Analytics in healthcare. *J Big Data*, *9*(1), 3. https://doi.org/10.1186/s40537-021-00553-4

Bradshaw, N. A. (2020). Florence Nightingale (1820-1910): An Unexpected Master of Data. *Patterns (N Y)*, *1*(2), 100036. https://doi.org/10.1016/j.patter.2020.100036

Brenner, M. J., Pandian, V., Milliren, C. E., Graham, D. A., Zaga, C., Morris, L. L., Bedwell, J. R., Das, P., Zhu, H., Lee, Y. A. J., Peltz, A., Chin, K., Schiff, B. A., Randall, D. M., Swords, C., French, D., Ward, E., Sweeney, J. M., Warrillow, S. J., . . . Roberson, D. W. (2020). Global Tracheostomy Collaborative: data-driven improvements in patient safety through multidisciplinary teamwork, standardisation, education, and patient partnership. *Br J Anaesth*, *125*(1), e104-e118. https://doi.org/10.1016/j.bja.2020.04.054

Cremer, F., Sheehan, B., Fortmann, M., Kia, A. N., Mullins, M., Murphy, F., & Materne, S. (2022). Cyber risk and cybersecurity: a systematic review of data availability. *Geneva Pap Risk Insur Issues Pract*, *47*(3), 698-736. https://doi.org/10.1057/s41288-022-00266-6

Dicuonzo, G., Galeone, G., Shini, M., & Massari, A. (2022). Towards the Use of Big Data in Healthcare: A Literature Review. *Healthcare (Basel)*, *10*(7). https://doi.org/10.3390/healthcare10071232

Hassan, M., Awan, F. M., Naz, A., deAndrés-Galiana, E. J., Alvarez, O., Cernea, A., Fernández-Brillet, L., Fernández-Martínez, J. L., & Kloczkowski, A. (2022). Innovations in Genomics and Big Data Analytics for Personalized Medicine and Health Care: A Review. *Int J Mol Sci*, *23*(9). https://doi.org/10.3390/ijms23094645

Hirsch, J., & Pitak-Arnnop, P. (2023). [Epidemiology and treatment of traumatic facial nerve palsy following skull base fractures : Results from the IBM MarketScan commercial database (2006-2019)]. *Unfallchirurgie (Heidelb)*. https://doi.org/10.1007/s00113-023-01367-0

Johnson, K. B., Wei, W. Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., Zhao, J., & Snowdon, J. L. (2021). Precision Medicine, AI, and the Future of Personalized Health Care. *Clin Transl Sci*, *14*(1), 86-93. https://doi.org/10.1111/cts.12884

Lorkowski, J., & Pokorski, M. (2022). Medical Records: A Historical Narrative. *Biomedicines*, *10*(10). https://doi.org/10.3390/biomedicines10102594

Ma, J., Johnson, E. A., & McCrory, B. (2023). Predicting risk factors for pediatric mortality in clinical trial research: A retrospective, cross-sectional study using a Healthcare Cost and Utilization Project database. *J Clin Transl Sci*, *7*(1), e211. https://doi.org/10.1017/cts.2023.634

Matabele, M. N., Cheng, C., Venkatesh, M., Barr, S., Ellefson, J., Beninati, M., Lobeck, I. N., & Puricelli, M. D. (2023). Perinatal airway management in neonatal goiter: A healthcare cost and utilization project (HCUP) kids' inpatient database analysis. *Int J Pediatr Otorhinolaryngol*, *175*, 111767. https://doi.org/10.1016/j.ijporl.2023.111767

Metcalfe, D., Zogg, C. K., Haut, E. R., Pawlik, T. M., Haider, A. H., & Perry, D. C. (2019). Data resource profile: State Inpatient Databases. *Int J Epidemiol*, *48*(6), 1742-1742h. https://doi.org/10.1093/ije/dyz117

National Institute of Standards and Technology. (2015). *NIST Big Data Interoperability Framework: Volume 1, Definitions* National Institute of Standards and Technology. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf

Osama, M., Ateya, A. A., Sayed, M. S., Hammad, M., Pławiak, P., Abd El-Latif, A. A., & Elsayed, R. A. (2023). Internet of Medical Things and Healthcare 4.0: Trends, Requirements, Challenges, and Research Directions. *Sensors (Basel)*, *23*(17). https://doi.org/10.3390/s23177435

Premo, H., Gordee, A., Lee, H. J., Scales, C. D., Moul, J. W., & Peterson, A. (2023). Disparities in Prostate Cancer Screening for Transgender Women: An Analysis of the MarketScan Database. *Urology*, *176*, 237-242. https://doi.org/10.1016/j.urology.2023.03.016

Saraswathula, A., Roy, S., Blythe, W. R., Gourin, C. G., & Boss, E. F. (2023). The Unrealized Potential of the Reg-ent ENT Clinical Data Registry. *JAMA Otolaryngol Head Neck Surg*, *149*(8), 659-661. https://doi.org/10.1001/jamaoto.2023.1389

Sarker, I. H. (2021). Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. *SN Comput Sci*, *2*(5), 377. https://doi.org/10.1007/s42979-021-00765-8

Schmalbach, C. E., Brereton, J., Bowman, C., & Denneny, J. C., 3rd. (2021). American Academy of Otolaryngology-Head and Neck Surgery/Foundation Reg-ent Registry: Purpose, Properties, and Priorities. *Otolaryngol Head Neck Surg*, *164*(5), 964-971. https://doi.org/10.1177/0194599820984135

Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. M. Z., Paul, R., Smriti, K., Naik, N., & Somani, B. K. (2022). The role of data science in healthcare advancements: applications, benefits, and future prospects. *Ir J Med Sci*, *191*(4), 1473-1483. https://doi.org/10.1007/s11845-021-02730-z

Uddin, M., & Syed-Abdul, S. (2020). Data Analytics and Applications of the Wearable Sensors in Healthcare: An Overview. *Sensors (Basel)*, *20*(5). https://doi.org/10.3390/s20051379

Wadhwa, H., Varshneya, K., Stienen, M. N., & Veeravagu, A. (2023). Do Epidural Steroid Injections Affect Outcomes and Costs in Cervical Degenerative Disease? A Retrospective MarketScan Database Analysis. *Global Spine J*, *13*(7), 1812-1820. https://doi.org/10.1177/21925682211050320

Xie, F., Beukelman, T., Sun, D., Yun, H., & Curtis, J. R. (2023). Identifying inpatient mortality in MarketScan claims data using machine learning. *Pharmacoepidemiol Drug Saf*, *32*(11), 1299-1305. https://doi.org/10.1002/pds.5658

Yogesh, M. J., & Karthikeyan, J. (2022). Health Informatics: Engaging Modern Healthcare Units: A Brief Overview. *Front Public Health*, *10*, 854688. https://doi.org/10.3389/fpubh.2022.854688