

Data Literacy and Data Science Literacy for Nurses: State of the Art Literature Review

Mikyoung LEE^{a,1}, Pankaj Vyas^b, Fabio D'AGOSTINO^c, Ann WIEBEN^d, Cynthia COVIAK^e, Margaret MULLEN-FORTINO^f, Suhyun PARK^g, Marisa SILEO^h, Sophia BROWNⁱ, Emiliane Nogueira DE SOUZA^j, Lisiane PRUINELLI^k, and Jethrone ROLE^l

^{a,1}Texas Woman's University, Dallas, TX, USA

^bSan Antonio Regional Hospital, Upland, CA, USA

^cSaint Camillus International University of Health Sciences, Rome, Italy

^dUniversity of Wisconsin-Madison, Madison, WI, USA

^eGrand Valley State University, Allendale, MI, USA

^fPenn Presbyterian Medical Center, Philadelphia, PA, USA

^gUniversity of Minnesota, Minneapolis, MN, USA

^hInterSystems, Cambridge, MA, USA

ⁱMemorial Sloan Kettering Cancer Center, New York, NY, USA

^jFederal University of Health Sciences of Porto Alegre, Porto Alegre, Brazil

^kUniversity of Florida, Gainesville, FL, USA

^lLoma Linda University Health, Loma Linda, CA, USA

ORCID ID: Mikyoun Lee <https://orcid.org/0000-0002-5316-3300>; Pankaj Vyas <https://orcid.org/0000-0001-8220-7408>; Fabio D'Agostino <https://orcid.org/0000-0002-4641-604X>; Ann Wieben <https://orcid.org/0000-0001-9275-0439>; Cynthia Coviak <https://orcid.org/0000-0001-8583-1986>; Suhyun Park <https://orcid.org/0000-0003-3414-4443>; Marisa Sileo <https://orcid.org/0009-0003-6685-7218>; Emiliane Nogueira de Souza <https://orcid.org/0000-0002-3873-4304>; Lisiane Pruinelli <https://orcid.org/0000-0002-1046-6037>

Abstract. In the rapidly evolving landscape of modern healthcare, nurses must proficiently navigate data utilization and grasp the principles of data science. Despite this urgency, nursing stakeholders currently do not fully understand the extent of data literacy or data science literacy they need to acquire. This paper aims to elucidate the distinctions between data literacy and data science literacy, offering insights into strategies for nurturing these competencies within nursing education, research, and practice. Through a state-of-the-art review of 22 articles and six healthcare industry resources, we identified a notable absence of comprehensive frameworks and assessment tools, highlighting key areas for future development.

Keywords. Data literacy, data science literacy, nursing

¹ Corresponding Author, Texas Woman's University College of Nursing, 5500 Southwestern Medical Ave. Dallas, TX, USA 75235; E-mail: mlee27@twu.edu

1. Introduction

In the dynamic landscape of modern healthcare, nurses, as frontline healthcare professionals, find themselves at the intersection of escalating data availability and its growing significance. It is imperative that nurses utilize data proficiently and grasp principles of data science. Therefore, there is a critical need for comprehensive education tailored to equip nurses with a nuanced understanding of both data literacy and data science literacy. Despite the essential role of data literacy and data science literacy in nursing practice and research, considerable ambiguity exists regarding the distinction between these literacies. This ambiguity extends to diverse nursing stakeholders, encompassing students, nurses, nurse practitioners, nurse administrators, nursing faculty, nurse informaticists, and nurse scientists and leaves a gap in the clarity of educational expectations in nursing. This review is to explore and synthesize multifaceted frameworks related to healthcare data management across academic education, practice, and industry. Additionally, the paper aims to elucidate the distinctions between data literacy and data science literacy, offering insights into the strategies needed to nurture these competencies within nursing education, research, and practice.

2. Methods

2.1. Study design and search strategy

We conducted a state-of-the-art literature review, which addresses emerging trends and synthesize insights from the latest literature and identify key areas that warrant further research [1]. Systematic literature searches were performed independently by three research team subgroups, with a focus on academia, clinical practice, and the healthcare industry. The searches were conducted from October to November 2023 using PubMed, ProQuest, Scopus, CINAHL, and Google. The search strategy, developed initially using MeSH terms and keywords in PubMed with Boolean operators, was then applied consistently across other databases. Key search terms included data science, data literacy, data competency, training, education, assessment, curriculum, clinical, practic*, health occupations, and health personnel. The search was limited to English-language literature published from August 2018 to August 2023. The initial search yielded a total of 330 unique articles for academia, 146 articles for clinical practice, and six industry resources.

2.2. Study selection and review process

We adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [2]. Included studies met the following criteria: 1) direct relevance to data literacy or data science literacy within the healthcare domain, 2) exploration of concepts, curricula, training, or assessment tools related to data literacy or data science literacy, and 3) publication in peer-reviewed journals, conference papers, books, reports, or materials. Article abstracts underwent initial screening by a pair of reviewers using Rayyan [3]. Following abstract screening, 40 and 21 articles were selected for full-text review by the academia and clinical practice subgroup, respectively. The first round of article reviews involved independent assessments by pairs of reviewers, while subsequent rounds engaged all reviewers until consensus on inclusion/exclusion was achieved. Ultimately, 22 articles (15 in academia and seven in clinical practice) and six industry resources were chosen for data

extraction. Data from each article were recorded in an Excel template, encompassing categories such as title, first author, year, study purpose, definitions of data/data science literacy, frameworks employed, domains of data/data science literacy, content and duration of training programs, study settings, target audiences, assessment tools for data literacy and data science literacy, and other outcomes. Findings were organized into definitions, components of frameworks, educational content, and assessment tools related to data/data science literacy across academia, clinical practice, and the healthcare industry.

3. Results

3.1. Definitions of Data Literacy and Data Science Literacy

No clear distinction in defining data literacy and data science literacy was found in the reviewed articles. Few articles provided refined definitions of data literacy, with some emphasizing healthcare-related literacies such as digital information literacy [4], genomic literacy [5], and public health informatics competencies [6]. Notable is the review of nursing informatics competency, the knowledge, skills, and attitudes for information use in nursing care, with a call for further research to establish core domains and assessment tools [7, 8]. Explicit definitions of data science literacy were lacking, but the articles presented varied definitions and scopes of data science. Data science was defined to encompass statistics, artificial intelligence (AI), machine learning (ML), database management, and systems engineering, with wide-ranging interpretations. Different terms like "Data Intelligence" [9] and "AI Literacy" [10] were introduced, emphasizing the transformation of data into value and covering principles like data governance, basic statistics, data visualization, and their impact on clinical processes.

3.2. Main and Sub-Components of the Data (Science) Literacy Frameworks

Various frameworks and models with diverse constructs and domains were cited, including data literacy framework [11], the European Commission digital competency framework [4], Data Intelligence Model [9], Data Science Curriculum Organizing Model [12], curricular categorizations [13], genomic nursing data science life cycle [14], the Health and Care Digital Capabilities Framework [15]. Scientific Reasoning Augmentation activities were categorized as problem identification, hypothesis generation, evidence generation, and communication [16].

3.3. Content of data (science) literacy education/training

Diverse healthcare professionals were targeted in data/data science training, with seven articles exclusively focusing on nurses and nursing students. An interdisciplinary approach was advocated in data science education, emphasizing consideration of the context with technical solutions driven by ML [17]. Data science content was proposed for nursing programs and professional training, including PhD nursing curricula using the Data Science Curriculum Organizing Model [12], genomic nursing education [14], general nursing curricula [18], critical care nurses [7], and school nurses [19].

Further broadening the scope, various experts shared data science course outlines designed for general healthcare professionals [15], biomedical researchers [20], and biomedical students [21]. Diverse modules were proposed, including data intelligence training modules [9], health data science competency courses [13], six data science programs [22], genomic literacy training with culturally appropriate communication

[5], and a comprehensive AI education approach, incorporating data governance, algorithmic decision-making, data visualization, and AI processes [10].

3.4. *Data (science) literacy assessment tools*

There was a notable absence of comprehensive data/data science literacy assessment tools. Swan [23] employed an AI competency questionnaire; however, information on psychometric properties was absent. Scientific Reasoning and Augmentation were measured using a decision scenario in a medical context [16]. A Digital Competency Tool assessed information processing, content creation, communication, safety, and problem-solving skills [4]. Public health informatics competencies were evaluated through four questions [6]. A Nursing Informatics Competency Assessment Tool, was used to examine the impact of informatics competency training [7].

3.5. *Data literacy and data science literacy resources in the healthcare industry*

The six industry web resources primarily aimed at enhancing data literacy for various organizations and personnel, only one explicitly targeting healthcare professionals. Notably, no resources were found for data science literacy. These tools offer competency-based frameworks for acquiring data literacy, presenting personas aligned with different data literacy levels (e.g., the categories Data Guru, Data Apprentice, Data Newcomer, and Data Avoider on qlik.com or Level 0-9 on healthcatalyst.com). These resources typically feature hands-on formats and/or online modules. While they are valuable for advancing data literacy based on need assessments, a notable gap exists in the form of assessment tools to evaluate acquired knowledge and training outcomes.

4. Discussion

This review revealed that the emergence of frameworks and models, indicating a growing interest in structuring data literacy and data science education potentially applicable to nurses. However, the absence of comprehensive assessment tools underscores a critical gap in evaluating training programs effectiveness and identifying areas for improvement. The industry's focus on data literacy suggests a recognition of its importance, but the dearth of resources for data science literacy targeting healthcare professionals, including nurses, highlights an area for future development. The lack of a clear distinction between data literacy and data science literacy highlights a need for standardized definitions and frameworks in the field.

5. Conclusions

This state-of-the-art review reveals diverse frameworks and models aimed at structuring data literacy and data science literacy. While data literacy is essential for understanding data concepts and interpreting data, data science literacy enables nurses to apply advanced analytical techniques to extract meaningful insights from complex datasets. The observed variations and paucity of comprehensive frameworks and related assessment tools underscore the need for future development. Future research and development efforts should prioritize addressing these gaps to better support nurses and other healthcare professionals in navigating the complexities of data utilization and the principles of data science in healthcare.

References

- [1] Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J*. 2009;26(2):91-108. doi:10.1111/j.1471-1842.2009.00848.x
- Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. Published 2021 Mar 29. doi:10.1136/bmj.n71
- [2] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan — a web and mobile app for systematic reviews. *Syst Rev*. 2016; 5:210. doi:10.1186/s13643-016-0384-4.
- [3] Shiferaw KB, Tilahun BC, Endehabtu BF. Healthcare providers' digital competency: a cross-sectional survey in a low-income country setting. *BMC Health Serv Res* 20, 2020:1021. doi:10.1186/s12913-020-05848-5
- [4] Ramos IN, Ramos KN, Ramos KS. Driving the precision medicine highway: community health workers and patient navigators. *J Transl Med* 2019;17:85. doi:10.1186/s12967-019-1826-2
- [5] McFarlane TD, Dixon BE, Grannis SJ, Gibson PJ. Public health informatics in local and state health agencies: An update from the public health workforce interests and needs survey. *J Public Health Manag Pract*. 2019;25 (2 Suppl):S67-S77. doi:10.1097/PHH.0000000000000918
- [6] Jouparinejad S, Foroughameri G, Khajouei R. et al. Improving the informatics competency of critical care nurses: results of an interventional study in the southeast of Iran. *BMC Med Inform Decis Mak*. 2020;20:220. doi:10.1186/s12911-020-01244-5
- [7] Kleib M, Chauvette A, Furlong K, et al. Approaches for defining and assessing nursing informatics competencies: a scoping review. *JBIM Evid Synth*. 2021;19(4):794-841. doi:10.11124/IBIES-20-00100
- [8] Gutiérrez-Aguado A, Curioso WH, Machicao JC, Eguia H. Strengthening capacities of multidisciplinary professionals to apply data science in public health: Experience of an international graduate diploma program in Peru. *Int J Med Inform*. 2023;169:104913. doi:10.1016/j.ijmedinf.2022.104913
- [9] Wiljer D, Hakim Z. Developing an artificial intelligence-enabled health care practice: Rewiring health care professions for better care. *J Med Imaging Radiat Sci*. 2019;50(4 Suppl 2):S8-S14. doi:10.1016/j.jmir.2019.09.010
- [10] Maybee C, Zilinski L. Data informed learning: A next phase data literacy framework for higher education. *Proceedings of the Association for Information Science and Technology*, 2016;52. doi:10.1002/pr2.2015.1450520100108
- [11] Shea KD, Brewer BB, Carrington JM, et al. A model to evaluate data science in nursing doctoral curricula. *Nurs Outlook*. 2019;67(1):39-48. doi:10.1016/j.outlook.2018.10.007
- [12] Beyene J, Harrar SW, Altaye M, et al. A Roadmap for building data science capacity for health discovery and innovation in Africa. *Front Public Health*. 2021;9:710961. doi:10.3389/fpubh.2021.710961
- [13] Dreisbach C, Koleck TA. The state of data science in genomic nursing. *Biol Res Nurs*. 2020;22(3):309-318. doi:10.1177/1099800420915991
- [14] Davies A, Hooley F, Eleftheriou I, et al. Applying co-design principles for the development of health education and workforce development. *Stud Health Technol Inform*. 2022;298:39-45. doi:10.3233/SHTI220904
- [15] Schmidt FM, Zottmann JM, Sailer M, et al. Statistical literacy and scientific reasoning & argumentation in physicians. *GMS J Med Educ*. 2021;38(4):Doc77. doi:10.3205/zma001473
- [16] Tolsgaard MG, Boscardin CK, Park YS, et al. The role of data science and machine learning in Health Professions Education: practical applications, theoretical contributions, and epistemic beliefs. *Adv Health Sci Educ Theory Pract*. 2020;25(5):1057-1086. doi:10.1007/s10459-020-10009-8
- [17] Foster M, Tasnim Z. Data science and graduate nursing education: A critical literature review. *Clin Nurse Spec*. 2020;34(3):124-131. doi:10.1097/NUR.0000000000000516
- [18] Bergren MD, Maughan ED. Data and information literacy: A fundamental nursing competency. *NASN Sch Nurse*. 2020;35(3):140-142. doi:10.1177/1942602X20913249
- [19] Heredia-Negron F, Alamo-Rodriguez N, Oyola-Velazquez L, et al. Evaluation of AIML + HDR-A course to enhance data science workforce capacity for Hispanic biomedical researchers. *Int J Environ Res Public Health*. 2023;20(3):2726. doi:10.3390/ijerph20032726
- [20] Moore JH, Boland MR, Camara PG, et al. Preparing next-generation scientists for biomedical big data: artificial intelligence approaches. *Per Med*. 2019;16(3):247-257. doi:10.2217/pme-2018-0145
- [21] Swan BA. Assessing the knowledge and attitudes of registered nurses about artificial intelligence in nursing and health Care. *Nursing Economics*. 2021;39(3):139-143.
- [22] Awad CS, Deng Y, Kwagyan J, et al. Summary of year-one effort of the RCMI consortium to enhance research capacity and diversity with data science. *Int J Environ Res Public Health*. 2022;20(1):279. doi:10.3390/ijerph20010279