Applying Artificial Intelligence to Electronic Health Record Data to Advance Symptom Phenotyping: A Brief Practical Guide

Melissa D. Pinto, PhD, RN, FAAN
Jerrold M. Jackson, PhD

May 31, 2025
DOI: 10.3912/OJIN.Vol30No02Man03

Article

Abstract

Symptom science has long been central to nursing practice, as symptoms are often the primary reason patients seek care. Nurses have historically relied on emerging research and clinical intuition to predict patient outcomes. Artificial intelligence (AI) and machine learning (ML) can now assist in developing phenotypes and may help nurses operationalize patients at risk. As AI technologies become more integrated into healthcare, the use of electronic health record (EHR) data for symptom phenotyping is emerging as a promising area for advancing care. However, nursing documentation remains underutilized in AI model development. This article provides a practical guide for nurses interested in using AI and ML to develop symptom phenotypes, covering key areas such as data availability, patient safety, model building, and commercialization. It highlights the potential of AI to accelerate symptom science, shorten diagnostic timelines, and improve patient outcomes by uncovering new treatment targets. While nurses typically lack formal training in AI, they possess valuable clinical expertise and are key collaborators in AI-driven research. This guide aims to empower nurses to engage in the development and implementation of AI solutions, ultimately contributing to more precise, individualized care and transformative advancements in healthcare.

Symptoms often serve as the primary reason that patients seek care.

Symptom science and management have long been central to the nursing discipline and the National Institutes of Nursing Research (Cashion & Grady, 2015). Symptoms often serve as the primary reason that patients seek care. Approximately 40-50% of primary care visits are driven by symptoms such as fatigue, dizziness, and headache, which lack a clear and unifying pathophysiology (Haller et al., 2015). These symptoms frequently represent early signs of emerging chronic diseases or acute health conditions (Agorastos & Chrousos, 2022).

For decades, nurses have been described as developing a "sixth sense" or "intuition" as they mature in their practice (Douw et al., 2015). This is especially so when a patient's condition appears ominous. This terminology typically refers to instances when nurses accurately predict outcomes despite clinical indicators being within normal limits. Such "intuition" may in fact be a distinct phenotype that could now be more reliably identified using artificial intelligence (AI) and machine learning (ML).

Better understanding of key data sources specific to nursing care and the parameters of change could inform symptom phenotyping.

As the healthcare sector increasingly adopts AI, symptom phenotyping applications are already emerging in the marketplace. These solutions address a variety of issues, including pre-symptomatic risk for sepsis, screening for radiological abnormalities, and calculating risk. Nursing documentation (e.g., nursing notes and flowsheets) is often underutilized as an important source of data for model training, refinement, and collaboration with nurses. Better understanding of key data sources specific to nursing care and the parameters of change could inform symptom phenotyping (Rossi et al., 2023). This article aims to: (1) provide background on the evolution of symptom phenotyping within nursing science, and (2) offer guidance for nurses interested in developing symptom phenotypes using AI and ML.

The Evolution of Chronic Disease Symptom Phenotyping and the Role of AI

Around 10 years ago, the National Institute of Health (NIH) convened an expert panel to advance symptom science through research on symptom clusters. The panel recommended a transformative approach to the assessment and management of multiple-occurring symptoms, with goals to improve patient outcomes and reduce healthcare utilization and costs

(Miaskowski et al., 2017). At that time, while machine learning and AI were recognized as valuable tools to advance symptom phenotyping, they were not yet fully accessible for the nursing discipline (Miaskowski et al., 2017).

Symptom phenotyping can be approached through "a priori" and "de novo" methods.

Symptom phenotyping can be approached through "a priori" and "de novo" methods. The former involves pre-specifying symptom clusters based on empirical evidence or clinical observations and using statistical approaches such as confirmatory factor analysis, path analysis, and structural equation modeling. The latter relies on qualitative methods and exploratory statistical techniques, including exploratory factor analysis and hierarchical cluster analysis.

This application of AI models has significantly reduced the translation gap from research findings to practice...

In 2025, we are now able to harness AI, combined with electronic health record (EHR) data, to uncover symptom phenotypes. This advancement accelerates symptom science by discovering new treatment targets and potentially shortening the time to diagnosis and treatment (Hens et al., 2023). While large data sets have been available for decades, only recently have multimodal, de-identified, and longitudinal data (both structured and unstructured) been made available on a national and soon-to-be global scale. Symptom science in nursing research is now entering a new era, expanding beyond traditional research to encompass the deployment of AI models into clinical and administrative workflows. This application of AI models has significantly reduced the translation gap from research findings to practice, which previously could take up to approximately 20 years, to just a few years or less (Aravazhi et al., 2025; Grady, 2025; Lenfant, 2023).

Intersection of Nursing Domain Expertise and AI Collaborators

Basic AI literacy, however, is becoming essential for successful collaboration...

Nurses, as expert clinicians, have more patient contact than any other healthcare providers. They possess intuitive knowledge of patient conditions and symptom patterns, which they may not always be able to fully articulate, but could potentially be uncovered through AI. Typically, traditional nursing PhD programs do not require courses on data mining or writing code to develop complex AI models. Basic AI literacy, however, is becoming essential for successful collaboration with colleagues in data science and software engineering.

Symptom Phenotyping Using AI and Machine Learning: Considerations for Nurses

To our knowledge, no published framework currently exists in the nursing literature to assist nurses to plan and collaborate with multidisciplinary colleagues on the use of AI and EHR data to create symptom phenotypes. In response, we have developed a set of practical questions to guide nurses from discovery to deployment of AI solutions. This guide, based on our extensive experience with industry-academic partnerships and AI startups in healthcare, addresses four key themes: (1) Data Availability and Security, (2) Patient Safety and Solution Design, (3) Model Building and Costs, and (4) AI Model Validation and Solution Commercialization. While scientists typically focus on traditional dissemination methods (e.g., publication and presentations), the success of AI-driven healthcare research may depend on its commercialization (Guindalini et al., 2021). Thus, we have included our guiding questions here in an effort to offer a usable framework for nurses to consider:

Data Availability & Data Security:

- What data are available, and in what format?
- Are the data centralized or decentralized across multiple repositories or institutions?
- Are the data harmonized (disparate sources brought together in a unified way) and/or standardized (data following the same format within a single dataset)?
- Who owns the data, and how can access be obtained?
- How are the data de-identified, and what limitations arise from the de-identification technique?
- How frequently are data fields updated, and does de-identified data change with each update?
- If nursing data are available on symptoms and are in an unstructured form, what potential analytic techniques are needed to transform data? What skillset is required among collaborators needed to complete analyses?

Patient Safety & Solution Design:

- Does the proposed project qualify as "human subjects" research?
- Is self-exemption possible, or does the data owner already have an IRB approval in place?
- What is the institutional policy on using AI applied to EHR data for algorithm development?

- Is an IRB approval still required if the data is de-identified?
- Will a human be involved in or near the loop, and what are the risks of fully autonomous versus human-involved AI systems?

Model Building & Costs:

- What are the expected compute costs, and who will cover them?
- Who are the best partners for model development (e.g., data scientists, software engineers), and are they internal or external to my/our institution?
  - If external, what permissions are required for data access?
  - What are the costs of collaborating with technical partners?

AI Model Validation & Solution Commercialization:

- How will the developed phenotypes be validated?
- How will I/we obtain an independent or holdout dataset for validation?
- What is my/our institutional process for deploying validated AI solutions into clinical workflows?
- What are the institutional policies on intellectual property (IP) and how will IP be protected?

Conclusion

Nurses are uniquely positioned to accelerate patient outcomes through phenotype discovery...

Nurses are uniquely positioned to accelerate patient outcomes through phenotype discovery, enabling more precise treatments and individualized care. However, few nurses have a background in data science or the business-related knowledge necessary for AI development and commercialization. Despite this, nurses are already actively involved in the development of AI solutions (Raymond et al., 2022). We hope this brief guide serves as a starting point, inspiring more nurses to engage in AI development and to contribute to solutions that can positively impact clinical practice and patient care.

Authors

Melissa D. Pinto, PhD, RN, FAAN
Email: Pinto.Melissa@mayo.edu
ORCID ID: https://orcid.org/0000-0002-1003-180X

Melissa D. Pinto is a nurse scientist and the Director of Nursing Research at Mayo Clinic Arizona. Dr. Pinto has more than a decade of experience in developing and testing electronic symptom management interventions. She and her team completed early discovery studies on long COVID, helping to define the prevalence rate, risk factors, and symptom phenotypes using the electronic health record.

Jerrold M. Jackson, PhD
Email: Jackson.Jerrold@mayo.edu

Jerrold M. Jackson is a seasoned data and technology executive with over 15 years of experience across healthcare, wellness, and financial services. He has led globally distributed and market facing data and technology efforts at large and mid-sized companies, as well as a variety of start-ups across multiple industries. Currently Dr. Jackson is Director, Partner Enablement at Mayo Clinic Platform. Prior to Mayo Clinic he was Chief Operating Officer at AIM7, Operating Partner at Synergistic Capital Partners, Advisor & Board Member with Synergistic Health Systems, Head/VP of Data & Machine Learning at Exos, and CEO at SHREDmill. Clinically, Dr. Jackson was a psychotherapist in Outpatient Psychiatry at Mount Sinai Medical Center.

References

Agorastos, A., & Chrousos, G. P. (2022). The neuroendocrinology of stress: The stress-related continuum of chronic disease development. *Molecular Psychiatry*, *27*(1), 502-513. https://doi.org/10.1038/s41380-021-01224-9

Aravazhi, P. S., Gunasekaran, P. , Benjamin, N. Z., Thai, A., Chandrasekar, Nikhil, K. K., Kolanu, D., Prajjwal, P., Tekuru, Y., Brito, L. V., & Inban, P. (2025). The integration of artificial intelligence into clinical medicine: Trends, challenges, and future directions. *Disease-a-Month, 71*(6), https://doi.org/10.1016/j.disamonth.2025.101882

Cashion, A. K., & Grady, P. A. (2015). The National Institutes of Health/National Institutes of Nursing Research intramural research program and the development of the National Institutes of Health symptom science model. *Nursing outlook*, *63*(4), 484-487. https://doi.org/10.1016/j.outlook.2015.03.001

Douw, G., Schoonhoven, L., Holwerda, T., Huisman-de Waal, G., van Zanten, A. R., van Achterberg, T., & van der Hoeven, J. G. (2015). Nurses' worry or concern and early recognition of deteriorating patients on general wards in acute care hospitals: A systematic review. *Critical care*, *19*, 1-11. https://doi.org/10.1186/s13054-015-0950-5

Grady. P. (2025, Feb. 19). *Partnering with OpenEvidence: A life-saving healthcare revolution*. Sequoia Capital. https://www.sequoiacap.com/article/partnering-with-openevidence-a-life-saving-healthcare-revolution/

Guindalini, C., Verreynne, M.-L., & Kastelle, T. (2021). Taking scientific inventions to market: Mapping the academic entrepreneurship ecosystem. *Technological Forecasting and Social Change*, *173*, 121144. https://doi.org/https://doi.org/10.1016/j.techfore.2021.121144

Haller, H., Cramer, H., Lauche, R., & Dobos, G. (2015). Somatoform disorders and medically unexplained symptoms in primary care. *Dtsch Arztebl Int*, *112*(16), 279-287. https://doi.org/10.3238/arztebl.2015.0279

Hens, D., Wyers, L., & Claeys, K. G. (2023). Validation of an Artificial Intelligence driven framework to automatically detect red flag symptoms in screening for rare diseases in electronic health records: Hereditary transthyretin amyloidosis polyneuropathy as a key example. *Journal of the Peripheral Nervous System*, *28*(1), 79-85. https://doi.org/10.1111/jns.12523

Lenfant, C. (2003). Clinical research to clinical practice—Lost in translation? *New England Journal of Medicine, 349*(9), 868–874. https://doi.org/10.1056/nejmsa035507

Miaskowski, C., Barsevick, A., Berger, A., Casagrande, R., Grady, P. A., Jacobsen, P., Kutner, J., Patrick, D., Zimmerman, L., & Xiao, C. (2017). Advancing symptom science through symptom cluster research: Expert panel proceedings and recommendations. *JNCI: Journal of the National Cancer Institute*, *109*(4), djw253. https://doi.org/10.1093/jnci/djw253

Raymond, L., Castonguay, A., Doyon, O., & Paré, G. (2022). Nurse practitioners' involvement and experience with AI-based health technologies: a systematic review. *Applied nursing research*, *66*, 151604. https://doi.org/10.1016/j.apnr.2022.151604

Rossi, L., Butler, S., Coakley, A., & Flanagan, J. (2023). Nursing knowledge captured in electronic health records. *International Journal of Nursing Knowledge*, *34*(1), 72-84. https://doi.org/10.1111/2047-3095.12365

Citation: Pinto, M.D., Jackson, J.M., (May 31, 2025) "Applying Artificial Intelligence to Electronic Health Record Data to Advance Symptom Phenotyping: A Brief Practical Guide" *OJIN: The Online Journal of Issues in Nursing* Vol. 30, No. 2, Manuscript 3.

Related Articles

ARTICLE May 31, 2025

An Ethics of Artificial Intelligence for Nursing

Jess Dillard-Wright PhD, MA, RN, CNM, FAAN; Jamie Smith, PhD, RN

ARTICLE May 31, 2025

Artificial Intelligence and Images Portraying Nurses Through the Decades

Janet Reed, PhD, RN, CMSRN; Tracy M. Dodson, PhD, RN; Amy B. Petrinec, PhD, RN; Delaney Tennant, BSN graduate; Jenna Chmelik, BSN, RN; Shawnna Cripple, BSN graduate

ARTICLE May 31, 2025

Artificial Intelligence in Nursing Practice: Decisional Support, Clinical Integration, and Future Directions

Garry Brydges, PhD, DNP, MBA, MHA, APRN, CRNA, ACNP-BC, FAANA, FAAN

ARTICLE May 31, 2025

Advancing Nursing Practice Through Artificial Intelligence: Unlocking Its Transformative Impact

Jennifer Shepherd, DNP, MHA, RN, NEA-BC, NPD-BC, CHPN; Amy McCarthy, DNP, RNC-MNN, NE-BC, CENP

ARTICLE May 31, 2025

Digital Defense Toolkit: Protecting Ourselves from Artificial Intelligence-Related Harms

Rae Walker, PhD, RN