The Golden Rule for AI Model Inputs of Nurse-Generated Data: *Do Not Assume*

Sarah C. Rossetti, RN, PhD[1,2], Rachel Y. Lee, RN, PhD[1], Kenrick Cato, PhD, RN, CPHIMS[3], David Albers, PhD[4], Shalmali Joshi, PhD[1], Varsha Varkhedi, BS[1], Amy Finnegan, PhD[5], Pinyue Vicky Wang, BBA[1], Victoria L. Tiase PhD, RN, NI-BC[6,7]

[1]Department of Biomedical Informatics, Columbia University, New York, NY; [2]School of Nursing, Columbia University, New York, NY; [3]University of Pennsylvania, Philadelphia, PA; [4]Biomedical Informatics, University of Colorado, Boulder, CO; [5]IntraHealth International, Chapel Hill, NC; [6]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT; [7]College of Nursing, University of Utah, Salt Lake City, UT

Corresponding Author:
Sarah Rossetti, RN, PhD, FACMI, FAMIA
Department of Biomedical Informatics, Columbia University
622 West 168th Street, PH20 WS-20
New York, NY 10032

Email: sac2125@cumc.columbia.edu
Phone: (781) 801-9211

Word Count: 2983
Keywords: nursing documentation, artificial intelligence (AI), electronic health records, data quality

Introduction

In hospital settings, registered nurses are the healthcare professionals who are at the bedside most frequently delivering and are primary contributors to the electronic health record (EHR).[1] As primary contributors to the EHR, nurses have been shown to document one flowsheet data point on average every minute during a clinical shift.[2] In addition to flowsheets, documentation by nurses also includes medication administration, narrative notes, and care team messages.[2] Nurses make many decisions over the course of a clinical shift, some that are protocol based and some that are not, and these decisions are reflected in their documentation.

Despite the volume and richness of nurse-generated data, these data remain poorly understood by non-nursing professionals, leading to few high-quality algorithms from non-nurse data scientists that effectively leverage nursing documentation. However, advanced artificial intelligence (AI) methods such as large language models offer new opportunities to mine and leverage large volumes of nursing data. As these methods continue to be optimized, we anticipate that data documented by nurses will be increasingly used in AI algorithms. These new uses of nursing data have substantial gains for bettering care but also carry new risks. Insufficient knowledge of nursing practice, nurse decision-making, and nursing workflows will risk inaccurate, misinterpreted, and undiscovered data signals. Recognizing how and why these data are generated by nurses is essential to developing AI models that are both clinically valid and accurately represent the complexities of real-world clinical workflows.

Foundationally, it is critical to understand that nurses do not collect data without a reason. Every data point exists for a specific purpose–either in response to a policy or requirements or as the result of a clinical judgment made by the nurse. The presence *or absence* of a data point cannot be separated from the nursing care process it reflects. Moreover, a nurse's primary responsibility is to provide care that returns a patient to a healthy state. While in some communities there is a misconception that nurses "check off" boxes in the EHR, these checkboxes are of course secondary to the nursing care provided. For example, much of the care documented is not only initiated and delivered by nurses but also frequently adapted in real time based on their clinical judgment to meet evolving patient needs. Any checkboxes are a matter of quality assurance and occasionally as a reminder. A more accurate view of nurses at the bedside delivering direct patient care shows they are not passive data collectors, but rather knowledge workers who are members of a dynamic and interprofessional care team that actively generate and interpret information to support clinical decision-making.[3,4] Some, but not all, of these data and information make their way into the EHR.

Knowledge-driven behaviors shape how nurses interact with the EHR, generating data that are not just administrative records, but reflections of underlying clinical decisions and care processes. Nursing documentation in an EHR is a living chronicle of the patient's care trajectory embedded within the care ecosystem for which the care team, including nurses, are active participants. For example, patterns of nursing documentation–such as unusually frequent assessments, entries made at uncommon times, or the use of optional fields to record additional context– have been demonstrated to be early behavioral signals of clinical concern by a nurse. These behavioral signals have been shown to predict patient deterioration up to 48 hours before changes in physiological indicators occur[5] and when used in an early warning score they influence significant decreases in mortality, sepsis and length of stay.[6]

Building effective AI tools from EHR data requires more than access to information. It demands clearly understanding how that information was generated. Traditionally, this understanding has been achieved by partnering directly with clinical subject matter experts, such as nurses, to select and validate the data that will serve as inputs in analytical models and to interpret the outputs, or results, in the context of the care that was delivered. However, with the increasing use of unsupervised learning in AI models, the expectations for manual oversight by expert clinicians when processing clinical data sets will likely change.

Recognizing the complexities in the evolving landscape described above, we propose the precept of "Do Not Assume" for researchers and developers working with nurse-generated data as inputs in AI models. We outline 12 heuristics as practical guidance for this precept describing common assumptions that, if left unchecked, can undermine the validity and clinical relevance of model outputs that use data documented by nurses. Some heuristics may also apply to data generated by other types of health professionals, yet our list is targeted for considerations necessary when using nurse-generated data. While the heuristics and examples below are not intended as an exhaustive list, based on our experience they are a robust starting point to build a useful crowdsourced community of resources to guide researchers and developers in thinking more critically and contextually about nurse-generated data.

1. **Do Not Assume that missing data equals missed care.**
   Nurses are busy, and their priority is safe, high-quality nursing care, often for many patients at once. A single patient's data are not an accurate representation of the totality of the situation. For example, if patient A has a code blue 1 minute after the nurse performs an action for patient B, this urgent code blue situation will not be observable from patient B's record and the nurse likely would not prioritize documenting any recent actions for patient B over saving the life of patient A; the nurse needs to be in patient A's room during a code blue, likely for the next hour or more. Alternatively, the nurse could forget to document or decide a particular event wasn't worth the effort to log into the EHR amid the long list of direct patient care needs and urgent requests they are responsible for meeting and the excessive burdens of the EHR.[7] Urgent situations occur frequently in the hospital and emergency department settings. Similarly, it is unwise to assume all fields were 'seen' by the nurse. Flowsheets are large and can rely on cascading functionality, meaning some fields may not be visible unless specific conditions are met, resulting in the busy nurse capturing the information in free-text instead.

2. **Do Not Assume protocols should be followed exactly as the logic states.** Clinical protocols are rough guidelines and cannot reflect every eventuality or situation due to the dynamic and complex nature of clinical care. Nurses may deviate from protocols not out of error, but out of necessity. Delivering safe, high-quality nursing care may require deviation from protocols due to the totality of a nurse's workload and higher clinical

priorities or unique patient situations that force decision points not explicitly addressed in the protocol.

3. **Do Not Assume that all data capture is equal**.
   Some data are captured because policy or requirements mandate them; others are recorded voluntarily because the nurse identified them as clinically relevant and important. Sometimes data capture happens for both reasons, but often this is not the case. Because of this, the presence of certain data points can signal heightened concern or nuanced judgment.

4. **Do Not Assume you understand the process that caused the data capture nor the temporal sequence of how data capture relates to actual care processes.**
   A nursing intervention may include administering a medication (e.g., administering Acetaminophen for a fever) or completing an action (e.g., removing an intravenous line that is no longer needed) or the absence of a medication (e.g., holding a blood pressure medication due to labile blood pressure readings that morning) or the purposeful lack of an action (e.g., not removing an intravenous line due to anticipated need during an upcoming procedure). A recorded physiological value may be an original assessment (e.g., oxygenation status prior to increasing supplemental oxygen) or a reassessment post-intervention (e.g., oxygenation status after increasing supplemental oxygen) and these may all be documented within the same hour. EHR data are a highly incomplete and almost entirely missing data source, and assessment and intervention fields are not consistently co-located within EHRs.[8]

5. **Do Not Assume you know the clinical procedures and protocols that apply to your population.**
   Protocols evolve over time and may differ by unit, clinician, institution or even within a given context in an institution, such as during severely busy versus more normal patient flow times.[9] Check the relevant clinical protocols with nurse subject matter experts and confirm if they are clinically relevant, used in the setting in which your data are generated, and whether they have changed during the years of your dataset.

6. **Do Not Assume that the same EHR has the same configurations and settings.**
   Even when institutions use the same EHR vendor, configurations can vary widely and can change monthly as the EHR software is updated. Health care facilities and their EHRs are highly dynamic, often staying the same for long periods followed by punctuated changes. For example, currently across the U.S., some health systems have nurses chart only abnormal findings (i.e., charting by exception) while others require nurses to explicitly confirm that all assessment findings fall “within defined limits” or “normal” range, often using a checkbox or shorthand entry. It’s important to ask- “Are nurses documenting by exception or within defined limits at a given health system?” And if latter, are the definitions for “within defined limits” consistent across the health system? Increasingly, some data may be streamed in from monitors and some data may be captured via ambient listening technology. These configuration differences may not be

visible in exported data, but they have significant implications for both modeling and interpretation.

7. **Do Not Assume that all values are reliable.**
Point-of-care measurements, such as bedside glucose checks or portable pulse oximetry, may offer needed speed and convenience to guide if further testing is needed. However, they can differ in accuracy, precision, and calibration when compared to lab-based measurements. Moreover, some point-of-care measurements are very accurate while others are known to be difficult to administer or have other problems that may severely impact measurement accuracy and reliability. It is essential to understand not only the differences between data sources but also the clinical rationale for choosing one method over another. For example, there are many point-of-care methods for measuring blood glucose, some of which are accurate while others are not, despite all being "blood glucose" measurements. EHR configurations also are not perfect, requiring nurses to use workarounds and therefore not using the system as configured, so that care can be delivered. For example, required documentation fields in the EHR may include "hard stops" that prevent the nurse from moving forward until a value is entered, even if the choices provided–or the actual clinical question–is not relevant to that patient. Without knowing the design and constraints of the documentation system, one risks misinterpreting the reliability and intent behind recorded values.

8. **Do Not Assume that all structured values make sense clinically or do not vary in time.**

   Evaluate distributions of data for each field and consult with nurse subject matter experts. Check for configuration issues (e.g., field is categorical but clinically the data is not categorical). Observe the distributions of data over time and look for stability and consistency as the data structures and documentation required change over time. Structured fields are not the same as clinical concepts mapped to standardized clinical terminology. In other words, just because data are structured does not mean they are clinically meaningful or appropriately configured. Some EHR fields may have categorical options that oversimplify complex clinical phenomena, for example categorizing nuanced assessment results into broad categories like mild, moderate, and severe. Before modeling, evaluate the distribution of values in each field and consult nurse subject matter experts to determine whether the data aligns with actual clinical practice.

9. **Do Not Assume a SQL query retrieved all the values needed for your modeling task.**
Nursing flowsheets are tricky–their complexity and volume make it easy to miss fields. For each clinical concept of interest in a dataset, investigate all the possible places in the EHR where the data can be entered, including across different flowsheet templates. Often the same clinical concept can be stored or visualized in many different back-end locations. Clarify which data fields you are interested in based on the modeling task. If data appear to be missing, check with nurses to confirm where in the EHR they enter those data and trace those fields from the user interface to the database tables. Verify

that all field names and values are spelled correctly, and search for common misspellings and abbreviations. Spot check your data pulls, comparing that the data extracted for a patient resembles what can be seen in the EHR system, noting any discrepancies or missing data that may be present.

10. **Do Not Assume the value of narrative data based on its length nor assume that low frequency fields are of low value.**
    Again, nursing flowsheets are challenging, and their complexity and volume make it easy to miss data. Nurses know this and take the extra time to link structured concepts in flowsheets using narrative comments. In nursing documentation, brevity does not mean irrelevance. Even a short narrative comment can convey critical clinical insight, while rarely used fields may capture important patient events or concerns. Given the complexity and density of nursing flowsheets, it is common for structured flowsheet fields to be overlooked or bypassed by the care team when reviewing data and nurses are aware of this. In response, nurses often use narrative comments to convey context, clarify assessment results, and to document care that doesn't fit into predefined structured fields. These data points may hold high value signals that are hidden and not found in structured data alone.

11. **Do Not Assume that information will be reliably captured in the same fields over time.**
    The EHR is a living, breathing entity. The combination of longitudinal configuration changes and the evolution of how users input information means that over time, the same data may be captured in different fields and in unexpected ways. Some data elements may fall into disuse entirely, and these processes can be slow or rapid. In addition to inspecting distributions of values for clinical information, the same approach should be used to understand how data capture for the same information changes over time. For example, clinical units are often mapped to a type of patient. At a given hospital, for example, everyone who works there will know that '10 North' is an oncology unit. It is computationally convenient to define oncology patients by determining who spent time on a bed in that unit. However, major change events do occur - for example, '10 North' may have had a pipe burst and require construction for 2 months, so for 2 months in your dataset the oncology patients may also have been on '10 South'. Another example is flowsheet template evolution. EHR analysts add new measures into templates to address evolving documentation needs. As removing outdated fields is structurally more challenging than adding new ones, and the usage frequency of certain EHR measures is uncertain, these outdated fields are frequently retained sometimes across different templates.[10] As a result, the same data types may appear in different fields.

12. **Do not assume that nursing data are statistically consistent over time, across different units, or across different institutions, nor assume that these data were extracted completely, correctly, or consistently.**

> It is important to develop methods to triangulate data consistency, check for changes in practice, and better interpret nursing and clinical data. Observe distributions of data elements specifically to: (i) investigate whether the distributions are logical and to identify outliers, errors, extraction mistakes, (ii) compare how these distributions change over time, noting that severe changes can be policy changes, nursing practice changes, clinical care practice changes, etc., (iii) compare gaps in documentation and measurement times according to clinical guidelines to catch errors and better understand nursing documentation practices, (iv) compare data types and data presence within and across clinical units as different clinical units can have very different documentation requirements, and (v) do not take this list as complete, but rather slice and re-analyze data in every way you can imagine by noting that robust results should not depend on how you partition data (e.g., via the concept of ergodicity, or averages over different partitions of data should be the same, if the processes generating the data are consistent.)

However, there is one item that is *fair game to assume* - if a nurse takes the time to capture information that is not otherwise required or often recorded, there is likely a very good reason and one worth exploring as a signal for an AI model input.

Conclusion

As AI models increasingly use EHR data as model inputs, it is imperative that researchers and developers move beyond a surface-level interpretation of nurse-generated data. These data are not incidental, they are shaped by nurses' clinical expertise, contextual judgment, institutional policies and the complex realities of patient care. Misinterpreting or oversimplifying them risks undermining the clinical validity of AI models and safety of their outputs. We acknowledge that while some of the heuristics and examples described above may also apply to data recorded by other types of health professionals, our list is specifically curated to convey the breadth of considerations necessary when using nurse-generated data.

We welcome input on, and additions to, the list of heuristics above in the form of comments to this pre-print publication to build a community-driven knowledge base for the AI community when using nurse-generated data. As stated above, the heuristics outlined here are not exhaustive but rather are intended to serve as a starting point to build a useful crowdsourced community of resources that are used by researchers and developers in thinking more critically and contextually about nurse-generated data. Recognizing the clinical expertise embedded in nursing documentation and carefully considering how to interpret those data are essential for developing AI models that are not only technically robust, but also clinically relevant and trustworthy.

Acknowledgements: The authors thank the many clinical nurse experts they have engaged with and learned from over the past decade to inform the defined heuristics.

Funding: *This project is supported by the Assistant Secretary for Technology Policy (ASTP) of the U.S. Department of Health and Human Services (HHS) under grant number and title for* *grant amount (90AX0042/01-02, Scalable, Shareable, and Computable Clinical Knowledge for AI-Based Processing of Hospital-Based Nursing Data, $998,903.00). This information or content and conclusions are those of the author and should not be construed as the official position or policy of, not should any endorsements be inferred by ASTP, HHS, or the U.S. Government.*

References

1. Jacques D, Will J, Dauterman D, Zavotsky KE, Delmore B, Doty GR, et al. Evaluating Nurses' Perceptions of Documentation in the Electronic Health Record: Multimethod Analysis. JMIR Nurs. 2025 Apr 28;8:e69651.
2. Collins S, Couture B, Kang M, Dykes P, Schnock K, Knaplund C, et al. Quantifying and Visualizing Nursing Flowsheet Documentation Burden in Acute and Critical Care. In: AMIA Annu Symp Proc. 2018.
3. Courtney KL, Alexander GL, Demiris G. Information technology from novice to expert: implementation implications. J Nurs Manag. 2008 Sep;16(6):692–9.
4. Hardiker NR, Dowding D, Dykes PC, Sermeus W. Reinterpreting the nursing record for an electronic context. Int J Med Inform. 2019 Jul;127:120–6.
5. Rossetti S, Knaplund C, Albers A, Dykes P, Kang M, Korach T, et al. Healthcare Process Modeling to Phenotype Clinician Behaviors for Exploiting the Signal Gain of Clinical Expertise (HPM-ExpertSignals): Development and Evaluation of a Conceptual Framework. J Am Med Informatics Assoc. 2021;28(6):1242–51.
6. Rossetti S, Dykes P, Knaplund C, Cho S, Withall J, Lowenthal G, et al. Real-time surveillance system for patient deterioration: a pragmatic cluster-randomized controlled trial. Nat Med 2025. 2025 Apr 2;1–8.
7. Levy DR, Withall JB, Mishuris RG, Tiase V, Diamond C, Douthit B, et al. Defining Documentation Burden (DocBurden) and Excessive DocBurden for All Health Professionals: A Scoping Review. Appl Clin Inform. 2024 Oct;15(5):898–913.
8. Hripcsak G, Albers DJ. Correlating electronic health record concepts with healthcare process events. J Am Med Inform Assoc. 2013 Dec;20(e2):e311-8.
9. Albers DJ, Elhadad N, Claassen J, Perotte R, Goldstein A, Hripcsak G. Estimating summary statistics for electronic health record laboratory data for use in high-throughput phenotyping algorithms. J Biomed Inform. 2018 Feb;78:87–101.
10. Wang P, Finnegan A, Yen P, Rossetti S. Mapping Documentation Burden: Analyzing Centrality and Clusters among Flowsheet Measures and Templates Through Network Analysis. In: AMIA Annual Symposium Proceedings. 2025.