

# **RNteract: Formulation of a Scalable and Extensible Logical Data Model of EHR Audit Logs for Temporal Data Mining**

Victoria L Tiase, Katherine A Sward, Julio C Facelli

Submitted to: JMIR Nursing  
on: December 28, 2023

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

*Table of Contents*

**Original Manuscript..... 5**  
**Supplementary Files..... 24**  
    Figures ..... 25  
        Figure 1..... 26



# RNteract: Formulation of a Scalable and Extensible Logical Data Model of EHR Audit Logs for Temporal Data Mining

Victoria L Tiase<sup>1\*</sup> PhD, RN; Katherine A Sward<sup>2\*</sup> PhD, RN; Julio C Facelli<sup>2\*</sup> PhD

<sup>1</sup>Department of Biomedical Informatics University of Utah Salt Lake City US

\*these authors contributed equally

## Corresponding Author:

Victoria L Tiase PhD, RN

## Abstract

**Background:** Increased workload, including workload related to electronic health record (EHR) documentation, is reported as a main contributor to nurse burnout, and adversely affects patient safety and nurse satisfaction. Although nurse burnout has been studied for decades, little has changed in the organization of clinical care. The measurement of nursing workload is not well understood. Traditional methods for workload analysis are either administrative measures (such as nurse-patient ratio) that do not represent actual nursing care, or are subjective and limited to snapshots of care (e.g., time-motion studies). Observing care, and testing workflow changes in real-time, can be obstructive to clinical care. An examination of EHR interactions through the use of EHR audit logs could provide a scalable, unobtrusive way to quantify nursing workload, at least to the extent that nursing work is represented in EHR documentation. EHR audit logs are extremely complex, however, and simple analytical methods can't discover complex temporal patterns, therefore requiring the use of state-of-the-art temporal data mining approaches. To effectively use these approaches it is necessary to structure the raw audit logs into a consistent and scalable data model that can be consumed by machine learning (ML) algorithms.

**Objective:** We aimed to conceptualize a data model for nurse-EHR interactions that would support the development of temporal ML models based on EHR audit log data.

**Methods:** We conducted a preliminary review of EHR audit logs to understand the types of nursing-specific data captured. Using concepts derived from literature and our previous experience studying temporal patterns in biomedical data, we formulated a data model that can describe the nurse-EHR interactions, nurse intrinsic and situational characteristics that may influence those interactions, and outcomes of relevance to nursing workload in a scalable and extensible manner.

**Results:** We described the data structure and concepts from EHR audit log data associated with nursing workload as a data model, that we name as RNteract. We conceptually demonstrated how using this data model could support temporal unsupervised machine learning and state-of-the-art artificial intelligence methods for predictive modeling.

**Conclusions:** The RNteract data model appears capable of supporting a variety AI-based systems (computational models), and should be generalizable to any type of EHR system or healthcare setting. Quantitatively identifying and analyzing temporal patterns of nurse-EHR interactions is foundational for developing interventions that support nursing documentation workload and address nurse burnout. Clinical Trial: Not Applicable

(JMIR Preprints 28/12/2023:55793)

DOI: <https://doi.org/10.2196/preprints.55793>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.  
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v  
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/preprint/55793>



## Original Manuscript

# RNteract: Formulation of a Scalable and Extensible Logical Data Model of EHR Audit Logs for Temporal Data Mining

## Abstract

**Background:** Increased workload, including workload related to electronic health record (EHR) documentation, is reported as a main contributor to nurse burnout, and adversely affects patient safety and nurse satisfaction. Traditional methods for workload analysis are either administrative measures (such as nurse-patient ratio) that do not represent actual nursing care or are subjective and limited to snapshots of care (e.g., time-motion studies). Observing care, and testing workflow changes in real-time, can be obstructive to clinical care. An examination of EHR interactions using EHR audit logs could provide a scalable, unobtrusive way to quantify nursing workload, at least to the extent that nursing work is represented in EHR documentation. EHR audit logs are extremely complex, however, and simple analytical methods cannot discover complex temporal patterns, therefore requiring the use of state-of-the-art temporal data mining approaches. To effectively use these approaches, it is necessary to structure the raw audit logs into a consistent and scalable logical data model that can be consumed by machine learning algorithms.

**Objective:** We aimed to conceptualize a logical data model for nurse-EHR interactions that would support the future development of temporal ML models based on EHR audit log data.

**Methods:** We conducted a preliminary review of EHR audit logs to understand the types of nursing-specific data captured. Using concepts derived from literature and our previous experience studying temporal patterns in biomedical data, we formulated a logical data model that can describe the nurse-EHR interactions, nurse intrinsic and situational characteristics that may influence those interactions, and outcomes of relevance to nursing workload in a scalable and extensible manner.

**Results:** We described the data structure and concepts from EHR audit log data associated with nursing workload as a logical data model, named RNteract. We conceptually demonstrated how using this logical data model could support temporal unsupervised machine learning and state-of-the-art artificial intelligence methods for predictive modeling.

**Conclusions:** The RNteract logical data model appears capable of supporting a variety AI-based systems and should be generalizable to any type of EHR system or healthcare setting. Quantitatively identifying and analyzing temporal patterns of nurse-EHR interactions is foundational for developing interventions that support nursing documentation workload and address nurse burnout.

**Keywords:** Burnout, professional; Nurses; Electronic Health Record; Data Modeling; Temporal Machine Learning

## Introduction

### Workload as contributor to burnout

Nursing workload - the amount of time, physical and cognitive effort needed to provide nursing care - is an ongoing concern in healthcare settings and can pose significant challenges for nurses.[1]

Increasing cost pressures and the lingering effects of COVID-19 have resulted in nurses taking care

of sicker patients than in the past.[2] Many healthcare facilities experience a shortage of qualified nursing staff, resulting from factors such as insufficient recruitment and retention efforts, or high turnover rates due to burnout.[2] There is also a higher demand for nurses given the aging population so that supply no longer meets demand.[2] As more nurses continue to exit the workforce, workload increases further for those who remain on the job.[3]

## **Measuring nursing workload**

Historically, the most used measure of nursing workload is the nurse-patient ratio,[4] referring to the number of patients assigned to a nurse during a specific shift or timeframe.[5] These ratios play a role in determining the amount of direct patient care that each nurse must provide, which directly impacts workload and the quality of care nurses can deliver. Research on patient ratios provides evidence connecting nurse staffing to patient outcomes and more specifically to patient safety.[4] However, establishing a safe and effective nurse-patient ratio is an ongoing challenge for healthcare organizations and policymakers.[6]

A major weakness to estimating nursing workload at this macro level is that it does not account for differences in patient illness or the amount of care needed by an individual patient, nor does it account for contextual and organizational characteristics that impact workload.[3] Additionally, recent studies found evidence that not all nurses, even within the same hospital unit, practice the same way, and therefore examining individual nurse practices becomes important in examining workload.[4] In addition to nurse-patient ratios, nursing workload has also been studied as a qualitative experience and estimated via time-motion studies, but the subjective and intrusive nature of these approaches suggests that more research should be done from a quantitative or mixed-methods perspective.[7,8]

## **EHR data as estimate of workload**

The healthcare industry has access to a massive amount of data that can be analyzed for trends,

patterns, and insights.[9] Data regarding clinical care are encompassed within the electronic health record (EHR). Researchers have leveraged EHR interactions to track clinical work activities and associated workload for physicians, for instance, and EHR interactions or use patterns were used to predict physician departures and burnout.[9-11] Nursing clinical care and documentation are distinct from physician care, but to date nursing workload has not been similarly evaluated from EHR data. Even though we know that nursing work extends beyond what is recorded, we posit that EHR documentation could be a reasonable proxy for nursing workload.

As part of regulatory requirements, healthcare organizations must record and track EHR activity at the user level, including logon attempts, what patient records are accessed, what documentation was entered, and the date and time of access.[12] These user-level metadata, stored as EHR audit logs, are an untapped resource that have potential to provide clinical insights.[13] Some nurse researchers have explored the suitability of EHR audit logs to understand documentation burden.[14] However, these metadata are complex, change over time, and varied in how they are aggregated into higher level measures.[15] To effectively use them, it is necessary to structure the raw audit logs as part of a consistent and scalable logical data model that can be consumed by artificial intelligence (AI) algorithms.

## Logical Data Models

A logical data model is a set of specifications that identify the primary data concepts and relationships between them; serving as a blueprint or template on how information is organized for analysis.[16] Machine learning models typically organize data as vectors. A vector, in this context, is a specific way to represent data as a matrix of values. In planning for machine learning, a logical data model can define the main categories of vectors that serve as model inputs or outputs, which can be implemented as a physical model consumable by any ML algorithm.[16] A logical data model is intentionally abstract, not constrained by either the data sources nor by the actual structure that will

store the data.<sup>16</sup> Ultimately, the aspects of the logical data model will inform the physical implementation of the model.[17]

## Objective

We posit that gathering and modeling sufficient data that are reliable, reproducible, generalizable, and that represent nursing contributions within the context of work activities and workload is achievable with data science-based research. We hypothesize that assessing nursing workload requires objective measurement and a standardization of data elements that represent clinical activities and other nursing workload influences that can be used as targets for modeling interventions at scale. This hypothesis is supported by the findings from the systematic review of studies using EHR audit logs to observe clinical activities by Rule et al.[10] Because EHR audit logs record all types of interactions with the health record, the logs may offer insights into how nurses interact with the EHR and the extent to which EHR interactions reflect workflow and workload. Our objective was to conceptualize a logical data model based on EHR audit log data as a first step towards analysis of nursing EHR documentation workflows. This is somewhat analogous to developing a conceptual framework before starting a traditional analysis. The physical implementation of the model will be described in subsequent work.

## Methods

This paper describes the development of a scalable and extensible logical data model to represent interactions between nurses and the electronic health records systems as recorded in EHR audit logs. We identified a general methodology for identifying concepts related to nursing workload and applied this methodology to EHR interactions as reflected in EHR audit logs. To develop this proposed conceptual framework, we began with a search of the literature, and expanded upon previous work that examined the workload of physicians using EHR data and researched components of nursing workload that can be extracted from other (not EHR) health system databases.[18] Two

experienced nurses iteratively refined the concepts and interactions until consensus was achieved. We conducted a preliminary review of the audit log from a commercial EHR vendor to confirm that the conceptualized model corresponded to generalizable structures for audit log data.

## Results

Using concepts derived from our previous work in studying temporal biomedical data patterns, we formulated a data structure that can describe nurse EHR interactions, nurse intrinsic and situational characteristics, and nurse outcomes of interest in a scalable and extensible manner.[19,20] We believe the selected features will allow for metadata aggregation into EHR use measures that can be used for a variety of nurse centric outcomes. Then, we conceptually instantiated the model with an analysis plan for quantitative study of the characteristics and expected outcomes associated with nurse EHR interactions using AI and temporal machine learning (ML) methods. Although our purpose is to focus on the EHR audit log data, we also provide details on additional data sources that will be needed to instantiate the logical data model. In the following sections, we discuss the EHR audit log metadata and the structure and components of the logical data model. We conceptualize how the model could be used to support data science methods employing AI and ML methods.

### EHR audit log metadata

In our review of the EHR audit logs, we found that nursing interactions with the EHR extend beyond clinical data input. Within the EHR audit logs, we were able to distinguish between data review and data input, the particular section of the EHR accessed (e.g., medication administration record, best practice advisories, or notes), and were able to identify workflow activities such as navigation between records. We focused on data specific to nurse users. In their current form, these data lack a hierarchical data structure and do not contain a taxonomy related to specific user or task types.

## Data model: concepts

The data model, which we named RNteract, contains elements that describe the tasks (NTask) performed by nurses (NType), for a specific type of patient or *patient panel* (NPanel), and that result in nurse-relevant outcomes (NOutcome). Each concept will be represented via a value set, for physical implementation of the model. The value sets can be defined based on the purpose or goals of the analysis. As an example, the NTask value set can describe a general type of task (like *medication tasks*) or can be more granular (specific steps of medication administration, for example) depending on the objectives of the model.

## Model structure: vectors

This logical data model is intended to support temporal ML models. These approaches are grounded in the concept of vectors or numerical arrays. A vector is a mathematically-based approach for expressing and organizing data in a pre-defined manner. Vectors in machine learning represent input data, including bias or weights. In the same way, output from a ML model can be represented as a vector. To perform a given ML task, the first step is to represent the input. For this logical data model, we provide a set of specifications and primary data structures with a focus on the NTask input.

## Model component: Nursing tasks (NTask)

Our initial focus will be on defining the nursing tasks (NTask). For each nurse,  $ID = k$ , we define a vector:  $\langle NTask(k, i) \rangle$ ,  $i=1$  to  $N$  as a vector of dimension  $N$ , where  $N$  is the number of time periods considered in the study. Any resolution (e.g., day, hour, 30-minute intervals, each minute, etc.) can be used for the time periods. The individual nurse performing the task is represented by  $k$ ; and the  $i$

component of the vector corresponds to the time interval under study.

The value representing the specific task is taken from a value set that describes the nursing tasks considered in the model and can be as general or specific as desired. For instance, if a value set for tasks has been defined broadly such that:

1 = No EHR interaction

2 = Read EHR data

3 = Input EHR data

and the nurse identified with ID = 55 has been reading the EHR records for the first time period, then does nothing with the EHR for the next three periods and writes a note in the EHR for the fifth time period, the corresponding NTask vector will be (55, 2, 1, 1, 1, 3).

The NTask vector will allow the classification of activity patterns in EHR data for any finite number of nurses, for an arbitrary set of tasks, and for any finite time resolution. Instantiation of the model would result in a set of vectors that can then be classified with temporal machine learning methods to model nurse activity trajectories and other patterns of nurse EHR interactions. Resulting nurse activity trajectories can be associated with quantitative descriptors of nurse characteristics (NType), the patient panel being cared for (NPanel), and outcomes or observations of nurse actions (NOutcomes), which are defined below.

## **Additional components**

### **Nurse type (NType)**

NType describes the nurse user within the EHR audit log data. For each nurse with ID = k, NType data element is a vector of dimension M in which the i component of the  $\langle \text{NType}(k, i) \rangle$ ,  $i=1$  to M, is an integer, real or categorical value defined by a value set that describes both intrinsic and situational characteristics of the nurse ID = k. The dimension of the vector, M, is defined by the number of characteristics we desire to model and to assign to a nurse.

Audit logs, by themselves, have only limited information about the nurse characteristics other than identification of individual users. From the EHR audit logs we can determine the location of where the EHR interaction occurred (inferred from the computer location), and consequently can infer based on most common location of documentation, characteristics such as whether the nurse primarily practices in, for example, a med-surg environment or an ICU. However, we will need to access data from other sources to incorporate other nurse characteristics. These data may be available from the EHR scheduling system or other auxiliary systems, like credentialing systems or human resources databases. Examples of a desired model value set, using these additional data sources, may be:

- 1: Nurse employment length; integer (e.g., number of years in the job, workplace).
- 2: Nurse highest degree; categorical (e.g., Licensed Practical Nurse, Registered Nurse (RN), Nurse Practitioner).
- 3: Average number of hours worked per week; integer. This is potentially derivable from the EHR audit log, e.g., based on first and last EHR interaction in a 24-hour period.
- 4: Nurse primary assignment; categorical (e.g., Operating Room (OR), Intensive Care, Medical/Surgical). This is potentially able to be inferred from the EHR audit log based on primary location of the computer used for documentation.

As an example, if nurse 55 has been in the organization for 5 years, is a RN, worked 48 hours per week and is primary assigned to the OR, the corresponding vector would be  $NType = (55, 5, RN, 3, OR)$ .

### **Patient panel (NPanel)**

NPanel is a vector to describe the patient context, for the tasks. The data element for the patient panel assigned to nurse with ID =  $k$  is a vector of dimension  $O$  in which the  $i$  element of  $\langle NPanel(k, i) \rangle$ ,  $i=1$  to  $O$  is an integer, real or categorical value defined by a value set that describes the

characteristics and the acuity of the patients in the panel. The dimension of the vector,  $O$ , is defined by the number of features used to describe the patient panel. For instance, this data element may accommodate the complexity level of the patients or the average length of stay of the patients. These data will not be available directly from EHR audit logs, but may be obtained from the admission, discharge, and transfer messages; a financial system; patient acuity; or other clinical sources. Examples of a possible value set for patient panel might be:

1. Admission from the Emergency Room; integer, number of patients admitted.
2. Admission from the Operating Room; integer, number of patients admitted.
3. Discharged to Home; integer, number of patients discharged.
4. Average length of stay; integer.

**An example for this component is if nurse 55 has a panel of 12 patients with 1 admission from the operating room, 2 discharges, and an average length of stay of 4, the corresponding vector would be  $N_{Panel} = (55, 0, 1, 2, 4)$**

### **Nursing outcome ( $N_{Outcome}$ )**

The  $N_{Outcome}$  element for the nurse with  $ID = k$  is a vector of dimension  $P$  in which the  $i$  element of  $\langle N_{Outcome}(k, i) \rangle$ ,  $i=1$  to  $P$  is an integer, real or categorical value defined by a value set that describes long term changes in the status of the nurse  $ID = k$  and  $P$  is the number of features included in the model to describe the nurse outcomes. This data element can accommodate events such as resignation, promotion, or salary increase. Other potential outcomes may include use of time off, patient safety events, and preventable harm. As with  $N_{Type}$ , these data may be available from the EHR or other ancillary systems. Examples of a value set may be:

- 1: Nurse salary increase; Boolean, YES, NO.
- 2: Nurse resignation; Boolean, YES, NO.
- 3: Nurse promotion; Boolean, YES, NO.

4: Nurse missed work; Integer, number of days that the nurse missed work.

For example, if nurse ID = 55 did not receive a salary increase, did not resign, was not promoted, but missed 6 days in the last month, the corresponding vector would be NOutcome = (55, NO, NO, NO, 6)

Such long-term outcomes will not be recorded in EHR audit logs and must be obtained from other data sources. However, audit logs may hint at intermediate outcomes and workflows, for example, workflow such as charting in a block of time (perhaps charting after the shift) versus charting throughout the shift. These intermediate outcomes may suggest hypotheses for further exploration, such as the extent to which charting in a block of time may reflect high workload.

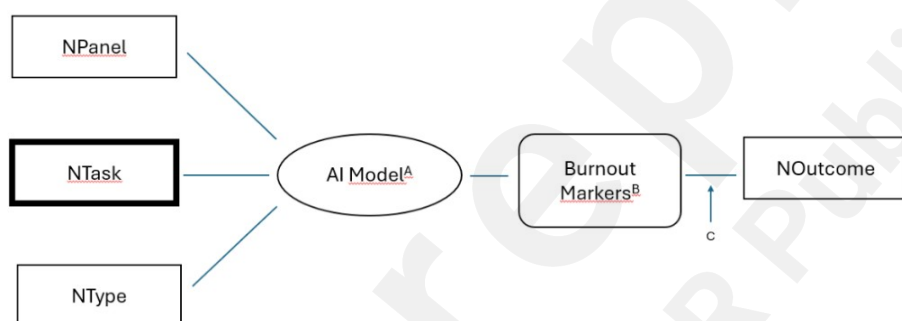
### **Extensibility/generalizability**

The definitions of NType, NPanel, and NOutcome are extensible and can accommodate any finite number of properties associated with any nurse. While we recognize that these properties may change over time, we chose to make NType, NPanel, and NOutcome independent of time because the variations in these categories are usually much slower than the variations of the time dependent NTask vector. If necessary, categorical variables can be transformed into integers by defining a table transformation. This makes NType, NPanel, and NOutcome amenable to any ML approach that may be considered.

### **Conceptual Instantiation for artificial intelligence and machine learning**

Using temporal unsupervised classification,[21-23] the  $\langle \text{NTask } (i, k) \rangle$  nurse EHR temporal interaction patterns can be classified through pattern recognition. The resulting archetypical activity patterns or clusters can be associated with any of the properties described in either NType, NPanel, or NOutcome, where the clusters can be described as enriched or depleted of any of the properties coded by these data elements. Using traditional statistics, the significance of the enrichments/depletions can be estimated.

The descriptive analysis of the NTasks clusters can be used to gain insights from interaction patterns which will allow for the development of predictive models for each cluster associated with the inputs for NTask, NType, and NPanel descriptors and outputs of interest described by NOutcome. As Figure 1 shows, this model allows for predictive modeling and is not constrained by the selection AI method. This intersection of nurse types, patient panels, and tasks, aggregated over time and across patient encounters, can be analyzed using a range of state-of-the-art AI techniques, for different purposes, revealing burnout markers such as charting, navigation, and searching patterns. Although we recognize burnout is not entirely measurable from EHR interactions, we assumed a set of burnout markers with concrete measurements that can be obtained from the user-centric logs. The logical data model also allows for the inclusion of other contributors to burnout that can be incorporated prior to describing the outcome.



**Figure 1.** RNteract Logical Data Model

A) Any state-of-the-art AI model can be accommodated; B) EHR-derived time patterns for charting, navigation, or searching; C) Allows for the inclusion of other contributors to burnout.

## Discussion

### Clinical Implications

AI-based systems have increasingly been incorporated into healthcare.[24] To date, little AI in healthcare research has considered real-world scenarios or real-world effect on outcomes.[25] Uses of AI in nursing span from virtual assistants to patient monitoring and predictive modeling.[26] AI

can also contribute by helping to streamline workflows and analyze vast amounts of data for evidence-based recommendations, helping nurses make more informed decisions about patient care. [26-28] Nurses have had only limited engagement in workflow or workload modeling efforts, despite the potential for AI-based systems to contribute to advanced, effective, efficient care; such as through more effective ways to access and organize information from EHRs.[25]

The metadata in audit logs can be complex and unstructured, making it difficult to interpret. Advanced analytics, like machine learning and data mining, are necessary to identify patterns and insights in this complex data, but to use these methods effectively, the data needs to be structured in a way that it is consumable by AI and ML applications.

The logical data model presented provides a structured framework for organizing the vast amount of data generated in audit logs. It helps in categorizing data elements and defining their relationships with a flow of logical reasoning, which is essential for understanding and using the data effectively. With a well-defined logical data model, it becomes easier to analyze audit log data by providing a clear representation of data flows and structures. We believe that this data model will provide a better understanding of the nursing related data elements and will assist others in using EHR audit log data to effectively model nursing workload or related outcomes of interest.

Prior to a physical implementation of the model (details not described here), the non-EHR audit log data sources may require additional representation. However, many of the other datasets have been described in the literature so our focus is to promote the broader use of EHR audit log data to understand nursing workload.[29,30]

## Limitations

The data model was developed based on literature and on the audit log of a single organization's electronic health record. The audit logs of other organizations using the same or different EHR may suggest additions or modifications to the model. Consultation with the nursing research community

may lead to further refinements of the model, to this end we envision subjecting the model to an open discussion in relevant research venues.

## Conclusion

Using concepts derived from our previous experience in studying temporal biomedical data patterns, we formulated a data model that can be used to describe nurse EHR interactions, intrinsic and situational characteristics of nurses and patient panels, and nurse outcomes of interest in a scalable and extensible manner. The definitions of NTask, NType, NPanel, and NOutcome are extensible and can accommodate any finite number of properties associated with a nurse. Through the use of the logical data model, we conceptualize how nurse EHR interactions could be studied using temporal unsupervised learning as well as any state-of-the-art AI methods. Future work may include extensions or modifications of the model as we test applicability to different organizations and different EHRs.

## Acknowledgements

All authors contributed to the conceptualization, model development, and writing of the manuscript. The research reported in this publication was supported in part by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UM1TR004409. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Conflicts of Interest

None declared.

## Abbreviations

AI: Artificial Intelligence

EHR: Electronic Health Record

ML: Machine Learning

OR: Operating Room

RN: Registered Nurse

## References

1. Ivziku D, Ferramosca FMP, Filomeno L, Gualandi R, De Maria M, Tartaglini D. Defining nursing workload predictors: A pilot study. *J Nurs Manag* 2022;30(2):473-481. doi:10.1111/jonm.13523
2. Fleisher LA, Schreiber M, Cardo D, Srinivasan A. Health Care Safety during the Pandemic and Beyond - Building a System That Ensures Resilience. *N Engl J Med* 2022;386(7):609-611. doi:10.1056/NEJMp2118285
3. Carayon P, Gurses AP. Nursing Workload and Patient Safety—A Human Factors

- Engineering Perspective. Agency for Healthcare Research and Quality (US); 2008. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21328758> [accessed September 15, 2023].
4. Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA* 2002;288(16):1987-93. doi:10.1001/jama.288.16.1987
  5. Driscoll A, Grant MJ, Carroll D, et al. The effect of nurse-to-patient ratios on nurse-sensitive patient outcomes in acute specialist units: a systematic review and meta-analysis. *European journal of cardiovascular nursing* 2018;17(1):6-22.
  6. Aiken LH, Lasater KB, Sloane DM, et al. Physician and nurse well-being and preferred interventions to address burnout in hospital practice: Factors associated with turnover, outcomes, and patient safety. *JAMA Health Forum* 2023;4(7):e231809. doi:10.1001/jamahealthforum.2023.1809
  7. Doleman G, De Leo A, Bloxsome D. The impact of pandemics on healthcare providers' workloads: A scoping review. *J Adv Nurs* 2023;doi:10.1111/jan.15690
  8. Higgins LW, Shovel JA, Bilderback AL, et al. Hospital nurses' work activity in a technology-rich environment: A triangulated quality improvement assessment. *J Nurs Care Qual* 2017;32(3):208-217. doi:10.1097/ncq.0000000000000237
  9. Sinsky CA, Rule A, Cohen G, et al. Metrics for assessing physician activity using electronic health record log data. *J Am Med Inform Assoc* 2020;27(4):639-643. doi:10.1093/jamia/ocz223
  10. Rule A, Melnick ER, Apathy NC. Using event logs to observe interactions with electronic health records: an updated scoping review shows increasing use of vendor-derived measures. *J Am Med Inform Assoc* 2022;30(1):144-154. doi:10.1093/jamia/ocac177
  11. Rule A, Chiang MF, Hribar MR. Using electronic health record audit logs to study clinical

- activity: a systematic review of aims, measures, and methods. *J Am Med Inform Assoc* 2020;27(3):480-490. doi:10.1093/jamia/ocz196
12. Auditboard. Understanding the HIPAA audit trail requirements: Essentials for compliance. 2023. <https://www.auditboard.com/blog/hipaa-audit-trail-requirements> [accessed September 5, 2023].
  13. Rose C, Thombley R, Noshad M, et al. Team is brain: leveraging EHR audit log data for new insights into acute care processes. *J Am Med Inform Assoc* 2022;30(1):8-15. doi:10.1093/jamia/ocac201
  14. Moy AJ, Hobensack M, Marshall K, et al. Understanding the perceived role of electronic health records and workflow fragmentation on clinician documentation burden in emergency departments. *J Am Med Inform Assoc* 2023;30(5):797-808.
  15. Rule A, Kannampallil T, Hribar MR, et al. Guidance For Reporting Analyses of Metadata on Electronic Health Record Use. *J Am Med Inform Assoc* 2023;doi:10.1093/jamia/ocad254
  16. McGilvray D. Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information. MIT Information Quality Industry Symposium. 2009. [http://mitiq.mit.edu/IQIS/Documents/CDOIQS\\_200977/Papers/01\\_02\\_T1D.pdf](http://mitiq.mit.edu/IQIS/Documents/CDOIQS_200977/Papers/01_02_T1D.pdf) [accessed September 4, 2023].
  17. Hughes R. R. Hughes, *Agile Data Warehousing for the Enterprise: A Guide for Solution Architects and Project Leaders*, Waltham, MA, USA:Morgan Kaufmann, 2016. <https://www.elsevier.com/books/agile-data-warehousing-for-the-enterprise/hughes/978-0-12-396464-9> [accessed September 4, 2023].
  18. Kannampallil T, Abraham J, Lou SS, Payne PRO. Conceptual considerations for using EHR-based activity logs to measure clinician burnout and its effects. *J Am Med Inform Assoc* 2021;28(5):1032-1037. doi:10.1093/jamia/ocaa305

19. Mistry S, Gouripeddi R, Raman V, Facelli JC. Sequential data mining of infection patterns as predictors for onset of type 1 diabetes in genetically at-risk individuals. *J Biomed Inform.* 2023;142:104385. doi:10.1016/j.jbi.2023.104385
20. Mistry S, Gouripeddi R, Facelli JC. Prioritization of infectious epitopes for translational investigation in type 1 diabetes etiology. *Journal of Autoimmunity* 2023;140:103115.
21. Zaki MJ. SPADE: An efficient algorithm for mining frequent sequences. *Machine learning* 2001;42(1):31-60.
22. Kour A. Sequential Rule Mining, Methods, and Techniques: A Review. *International Journal of Computational Intelligence Research* 2017;13(7):1709-1715.
23. Lo D, Khoo S-C, Wong L. Non-redundant sequential rules—Theory and algorithm. *Information Systems* 2009;34(4):438-453. doi:https://doi.org/10.1016/j.is.2009.01.002
24. Lee D, Yoon SN. Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. *International Journal of Environmental Research and Public Health* 2021;18(1):271.
25. Pailaha AD. The impact and issues of artificial intelligence in nursing science and healthcare settings. *SAGE Open Nursing* 2023;9:23779608231196847. doi:10.1177/23779608231196847
26. Yelne S, Chaudhary M, Dod K, Sayyad A, Sharma R. Harnessing the power of AI: A comprehensive review of its impact and challenges in nursing science and healthcare. *Cureus* 2023;15(11)
27. Schwartz JM, Moy AJ, Rossetti SC, Elhadad N, Cato KD. Clinician involvement in research on machine learning–based predictive clinical decision support for the hospital setting: A scoping review. *J Am Med Inform Assoc* 2021;28(3):653-663.
28. Korach ZT, Cato KD, Collins SA, et al. Unsupervised machine learning of topics documented by nurses about hospitalized patients prior to a rapid-response event. *Applied*

Clinical Informatics 2019;10(05):952-963.

29. Effken JA, Brewer BB, Patil A, Lamb GS, Verran JA, Carley KM. Using computational modeling to transform nursing data into actionable information. *Journal of biomedical informatics*. 2003 Aug 1;36(4-5):351-61.
30. Lindberg DS, Prosperi M, Bjarnadottir RI, Thomas J, Crane M, Chen Z, Shear K, Solberg LM, Snigurska UA, Wu Y, Xia Y. Identification of important factors in an inpatient fall risk prediction model to improve the quality of care using EHR and electronic administrative data: a machine-learning approach. *International journal of medical informatics*. 2020 Nov 1;143:104272.

## Supplementary Files

## Figures

RNteract Logical Data Model A) Any state-of-the-art AI model can be accommodated; B) EHR-derived time patterns for charting, navigation, or searching; C) Allows for the inclusion of other contributors to burnout.

