Investigating the (mis)match between electronic health records and actual nursing work: An observational study

Peer-reviewed author version

VANTHIENEN, An; MARTIN, Niels & DEPAIRE, Benoit (2026) Investigating the (mis)match between electronic health records and actual nursing work: An observational study. In: International journal of nursing studies, 175 (Art N° 105309).

DOI: 10.1016/j.ijnurstu.2025.105309
Handle: http://hdl.handle.net/1942/48082

INVESTIGATING THE (MIS)MATCH BETWEEN ELECTRONIC HEALTH RECORDS AND ACTUAL NURSING WORK: AN OBSERVATIONAL STUDY

An Vanthienen[a,b], Niels Martin[a,b] and Benoît Depaire[a,b]

[a]*UHasselt - Hasselt University, Faculty of Business Economics, Agoralaan, Diepenbeek, 3590, Belgium*
[b]*UHasselt - Hasselt University, Digital Future Lab, Wetenschapspark 2, Diepenbeek, 3590, Belgium*

ARTICLE INFO

Keywords:
Process mining
Healthcare
Electronic health records
Data quality
Observational study

ABSTRACT

Purpose: The process mining research domain uses process execution data to gain insights into work processes and has been applied in a wide variety of domains, including healthcare. The extensive use of electronic health record systems has made the data they capture a common input for process mining, yet data quality issues persist. While these issues are recognized in literature, empirical work examining the (mis)match between nursing interventions and their electronic health record registrations remains scarce. This study addresses this gap through an observational study. **Methods**: A cross-sectional observational study was carried out between February 23 and April 10, 2024, covering 119.75 hours of observation in the urology and neurology wards of a Belgian hospital. Data were collected on both nursing interventions and their electronic health record registrations. **Results**: The analysis revealed several mismatches between electronic health record registrations and actual nursing work: (i) 20.34% of all observed intervention types were never recorded in the EHR, (ii) only 23.32% of registered interventions were recorded without a time gap between execution and registration, and (iii) there is not always a one-to-one relationship between interventions and registrations. **Conclusion**: This study underscores the importance of thorough data quality assessment when using routinely collected data for research or analysis. Beyond assessing the data itself, it highlights the need to understand real-world work processes and how data is recorded in supporting systems. Such insights enable the anticipation and potentially the mitigation of data quality issues prior to their actual use. These efforts are essential to determine how accurately the available data reflects real-world practices and thereby how trustworthy any conclusions based on this data can be.

1. Introduction

Currently, hospitals are challenged to provide high-quality care while facing limited resources and increased care needs (Stalpers, Schoonhoven, Dall'Ora, Ball and Griffiths, 2025; Munoz-Gama, Martin, Fernandez-Llatas, Johnson, Sepúlveda, Helm, Galvez-Yanjari, Rojas, Martinez-Millana, Aloini et al., 2022). In addition, a key struggle in many countries is chronic nurse understaffing (Stalpers et al., 2025; Van den Heede, Balcaen, Bouckaert, Bruyneel, Cornelis, Sermeus and Van de Voorde, 2023). Given the context of these tight budgets and the significant shortages in nursing staff, hospitals are looking into ways to improve their processes (i.e. interrelated sets of activities, decisions and events with a particular goal (Dumas, La Rosa, Mendling, Reijers et al., 2013)) as they play a central role in a healthcare organization's daily operations (Martin, De Weerdt, Fernández-Llatas, Gal, Gatta, Ibáñez, Johnson, Mannhardt, Marco-Ruiz, Mertens et al., 2020).

In order to identify areas in which a work process can improve, it is crucial to have a good understanding of how the process is executed in reality. In this respect, process mining offers data-driven methods that generate insights into how work processes are executed based on digital traces recorded in information systems (van der Aalst, 2016). The generated process-related insights can relate, among others, to the order of activities in a process, the extent to which the real-life execution of a process conforms to a normative process flow, and the performance of a process (van der Aalst, 2022). Because of key advantages such as its ability to enhance the transparency of work processes, process mining has already been applied in a wide variety of domains, of which healthcare is a prominent one (Reinkemeyer, 2020; Zerbino, Stefanini and Aloini, 2021; De Roock and Martin, 2022). In healthcare, process mining has most commonly been used to analyze patient care pathways across different settings (Rojas, Munoz-Gama, Sepúlveda and Capurro, 2016) such as oncology (Baker, Dunwoodie, Jones, Newsham, Johnson, Price, Wolstenholme, Leal, McGinley, Twelves et al., 2017), stroke (Noshad, Rose and Chen, 2022), emergency care (Benevento, Dixit, Sani, Aloini and van der Aalst, 2019) and dental care (Fox, Whelton, Johnson and Aggarwal, 2023). Other use cases include assessing the adherence of a process to clinical guidelines (Huang, Dong, Ji, Yin and Duan, 2015; Rinner, Helm, Dunkl, Kittler and Rinderle-Ma, 2018), analyzing the time-related performance of a care process (Cho, Song, Park, Yeom, Wang and Choi, 2020; Yoo, Cho, Kim, Kim, Sim, Yoo, Hwang and Song, 2016), making process-related predictions (Cremerius, König, Warmuth and Weske, 2021; Jonk, Schaller, Netzer, Pfeifer, Ammenwerth and Hackl), and making process-related comparisons (Partington, Wynn, Suriadi, Ouyang and Karnon, 2015; Stefanini, Aloini, Benevento, Dulmin and Mininno, 2018). By providing in-depth data-driven insights into processes, process mining can contribute to support evidence-based process analysis and improvement in healthcare (Munoz-Gama et al., 2022).

Despite the promise of process mining, its success heavily depends on the quality of the data it relies on (Munoz-Gama et al., 2022). Electronic health record systems are increasingly used to support the execution of work processes in hospitals (Adler-Milstein, DesRoches, Kralovec, Foster, Worzala, Charles, Searcy and Jha, 2015; Jianxun, Arkorful and Shuliang, 2021). These information systems primarily aim to set care objectives, plan care, document care delivery and assess outcomes (Häyrinen, Saranto and Nykänen, 2008). Given the prominent use of electronic health record systems during the operational execution of work processes in hospitals, data from their underlying database is often used in process mining case studies. However, existing literature has also widely recognized the presence of data quality issues relevant to process mining in healthcare (Fox, Aggarwal, Whelton and Johnson, 2018; Munoz-Gama et al., 2022; Vanbrabant, Martin, Ramaekers and Braekers, 2019). For example, when nurses postpone the electronic health record registration of activities they have performed until a quieter moment, the timestamps of the registration will no longer correspond to the time at which the activity has been performed, posing challenges for process mining as it heavily relies on timestamp information (Vanbrabant et al., 2019). Other examples of data quality issues highlighted in literature are missing events (i.e. events that took place, but have not been registered in the electronic health records) and imprecise resource data (i.e. the specific nurse performing a task is not registered in the electronic health records) (Mans, van der Aalst and Vanwersch, 2015). The presence of data quality issues raises questions about the suitability of electronic health record data as a basis to study how work is actually performed in care processes as this data source is potentially an incomplete and/or inaccurate reflection of reality.

Until now, empirical work shedding light on the (mis)match between nursing interventions (i.e. the activities that nurses perform in a work process) on the one hand and registrations in the electronic health record on the other hand is limited. Existing literature primarily focuses on stating that data quality issues are present together with approaches on how to identify and/or handle these issues in an event log (Bose, Mans and Van Der Aalst, 2013; Mans et al., 2015; Suriadi, Andrews, ter Hofstede and Wynn, 2017; Vanbrabant et al., 2019). However, no systematic consideration is given to observing how work is being performed on the job and how employees interact with the information system at hand. This paper complements the body of literature by giving explicit consideration to how care is delivered, thus focusing on the care process while giving explicit attention to the interaction with the information system. By observing nursing interventions as well as the electronic health record registrations they perform, this paper provides a unique perspective on the (mis)match between electronic health record registrations and actual nursing work.

This paper complements earlier work by Fore, Islim and Shever (2019), who conducted an observational study to assess whether electronic health record data were sufficient to estimate nursing costs. While their study focused on measuring the time required to complete nursing tasks for cost calculation purposes, our study investigates whether electronic health record data are suitable as input for process mining applications, which aim to analyze and improve healthcare processes. While cost estimation mainly relies on understanding activity durations, process mining demands accurate reconstruction of the order and timing of nursing interventions. Therefore, rather than focusing on the duration of nursing interventions, we examined the extent to which electronic health record registrations accurately reflect the actual timing and sequence of nursing interventions, identifying potential mismatches that could impact the reliability of process mining results.

2. Methodology

The goal of this study is to investigate the (mis)match between the performed nursing interventions and their registrations in the electronic health record. To this end, an observational study with nurses as participants has been conducted at two distinct wards of a Belgian hospital.

2.1. Study Design

A cross-sectional observational study was conducted in which nurses of a hospital ward were observed for a limited time period. The observer maintained a non-participatory stance, allowing nurses to perform their interventions under normal circumstances. Nevertheless, occasional inquiries were necessary to clarify data entries in the electronic health record or specific interventions performed in a patient room (Mann, 2013). Information about both the execution of nursing interventions and the registration of data in the electronic health record were collected. Hence, this study entirely builds upon observational data (i.e. observed nursing interventions and observed data registrations); no data have been extracted from the hospital's information systems.

Besides the observations, an in-depth discussion with a nurse provided insights into the general organization of a shift and the typical nursing interventions and responsibilities. These insights guided data interpretation during the analysis stage.

2.2. Setting & participants

Observations were conducted at two distinct acute care wards of a Belgian hospital: an urology ward with 19 patient rooms (10 single rooms and 9 double rooms) and a neurology ward with 18 patient rooms (6 single rooms, 10 double rooms and 2 double rooms dedicated to stroke patients).

All participants were nurses employed at these hospital wards. Nurses working during the observed shifts were invited to participate, after which one nurse was selected for observation during each shift. The researcher documented the interventions and data entries performed by the observed nurse throughout the entire shift. To ensure anonymity, no personal information about the observed nurse was collected. Each nurse could only be observed once throughout the study to ensure that behavioral differences between nurses were captured.

2.3. Data collection & analysis

During each observed shift, observational data were recorded in two different datasets referred to as the Interventions and Registrations dataset. The Interventions dataset contains data about the performed nursing interventions. The Registrations dataset encompasses data about the observed entries made by nurses in the electronic health records.

Each record in the *Interventions dataset* corresponds to the execution of a single nursing intervention. For each intervention, data were collected on several aspects:

- The ward where the intervention took place;
- The type of shift (morning shift, afternoon shift or night shift) during which the intervention was performed;
- The type of intervention carried out by the nurse (e.g. measurement of vital signs, garbage disposal, consultation of the electronic health records, checking urinary catheter, ...);
- A timestamp, indicating the date and time at which the intervention was executed;
- The link with the associated record in the Registrations dataset (if a link with a registration could be observed);
- If necessary, notes could be added to clarify the record.

The *Registrations dataset* captures each observed electronic health record registration that the nurse performed. For each observed registration, the following data were recorded:

- A timestamp, indicating the date and time at which the nurse made the registration;
- The type of registration made into the electronic health records (e.g. registration of vital signs, medication administration, medication preparation, assisting in activities of daily living, ...);
- A marker indicating whether the registration was observed as a multi-registration, defined as the simultaneous registration of interventions of the same or different types for one or more patients;
- The link with the associated record in the Interventions dataset (if this could be observed).

Observations were conducted during different types of nursing shifts (morning, afternoon, night) and throughout the entire week to cater for potential differences.

To standardize the different nursing interventions that have been observed, the free-text observational data were mapped to the Nursing Intervention Classification. The Nursing Intervention Classification is a comprehensive, research-based and standardized classification of interventions performed by nurses. The classification offers a broad spectrum of interventions, ranging from direct patient care to indirect care such as administrative functions and supply chain management. The Nursing Intervention Classification taxonomy contains three levels to identify nursing activities: domains, which form the highest level of classification, classes and interventions. A nursing intervention is described as: *"Any treatment, based upon clinical judgment and knowledge, that a nurse performes to enhance patient/client outcomes. Nursing interventions include both direct and indirect care; those aimed at individuals, families and the community; and those for nurse-initiated, physian-initiated, and other provider-initiated treatments."* (Butcher, Bulechek, Dochterman and Wagner, 2018, p.48).

In this study, mapping was performed at the intervention level rather than the activity level. This choice reflected the level of granularity captured during the observations, which aligned more closely with the Nursing Intervention Classification intervention categories than with the more fine-grained activity definitions. This approach preserved the essential content of the observations while allowing the data to be clustered and analyzed in a standardized manner. For example, the observed intervention of wound care after surgery was mapped to the Nursing Intervention Classification intervention 'Wound Care', which encompasses a series of activities such as cleansing, dressing changes and inspection of the wound (Butcher et al., 2018). Moreover, the Nursing Intervention Classification is a frequently used classification in nursing literature, which allowed clustering different observed nursing interventions in order to guide data analysis and make results interpretable (Fennelly, Grogan, Reed and Hardiker, 2021).

After this mapping, both datasets were analyzed to discern patterns and relationships that shed light on the correspondence between performed nursing interventions and electronic health record registrations. From a practical perspective, the data analysis was performed using `R`[1], which is a programming language with extensive functionalities for data manipulation and statistical analysis. The key packages that were used are `dplyr` for data manipulation and summarization, and `ggplot2` for data visualization.

2.4. Ethical considerations

The study was conducted in accordance with all relevant ethical guidelines and regulations. Ethical approval was obtained from the ethics committees of both the researcher's academic institution (reference number CME2023/60) and the hospital (reference number VT2023-13) where the study took place. Participants were fully informed about the study's purpose and written informed consent was obtained prior to data collection.

3. Results

This section outlines the key results. These results emerged from analyzing the data of observations conducted between February 23, 2024 and April 10, 2024. Each shift lasted 8.5 hours (or 9.25 hours in case of a night shift). After a total of 119.75 hours of observation, data saturation was reached as no significant new information was collected. Appendix A contains the schedule of observed shifts.

3.1. Typical structure of a shift

Each shift starts with a *handover* where nurses review patient records and receive handover reports from their colleagues. During the afternoon and night shifts, the handover is followed by preparing the *medication* for the day. The afternoon shift gets the oral medication ready, whereas the night shift prepares intravenous drugs. Next, nurses *tour* the ward to administer medication and measure patients' vital signs. They also register their performed interventions in the electronic health records using mobile carts containing a laptop. In case of a morning shift or an afternoon shift, nurses subsequently make a *second tour* to assist people with their activities of daily living. After the second tour, nurses focus on the registration of all performed activities of daily living interventions in the *electronic health records* and update patient files. Most of the interventions performed to support activities of daily living are entered into a personalized activity plan that contains the individual care needs of a particular patient and consists of specific interventions to be carried out for the patient at hand. Following these entries, nurses check if patients are scheduled for examinations or surgery and if specimens need to be obtained. Afterwards, nurses attend patients in need of help,

[1]https://www.r-project.org

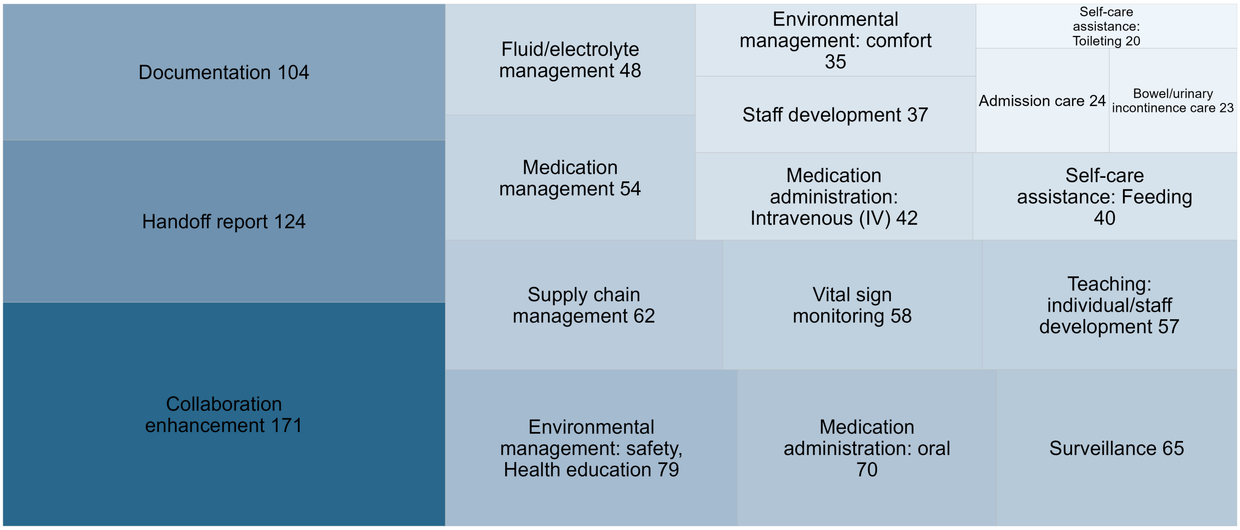

Figure 1: Treemap of nursing interventions that occurred more than 20 times

ensure specimens are collected and patients are prepared for surgery and examinations. Towards the end of the shift, nurses review patient files again and document noteworthy events. Shifts always end with a *handover* to the incoming colleagues of the next shift.

3.2. Overview of the Interventions and Registrations datasets

During the observations, nurses performed 1,435 interventions, each resulting in a record in the Interventions dataset. Likewise, nurses made 393 entries in the electronic health records, resulting in 393 records in the Registrations dataset.

3.2.1. Overview of the Interventions dataset

A total of 59 distinct nursing interventions are observed. Figure 1 presents a treemap highlighting interventions that occurred at least 20 times, with this threshold chosen to ensure the readability of the visualization. For the sake of completeness, all interventions and their frequency of occurrence are summarized in Appendix B. Some interventions are performed substantially more than others: the 10 most frequently observed interventions account for 60.60% of the total number of observed interventions (869 out of 1,435 observed interventions). The most frequently observed intervention is ‘Collaboration enhancement’, highlighting the frequent collaboration among nurses and other healthcare workers (171 instances). Following are the interventions ‘Handoff report’ (124 instances) and ‘Documentation’ (104 instances). These top three interventions, all categorized as indirect care in nursing literature, involve interventions essential to high-qualitative care without involving direct patient contact (Alghamdi, 2016). Interestingly, the highest-ranking direct care intervention (‘Medication administration: oral’) was observed 70 times and was only ranked fifth.

Upon analyzing the frequency of interventions across the two hospital wards, some notable variations were observed. Figure 2 highlights the ten interventions with the largest differences in relative frequency of occurrence. Two interventions stand out with particularly pronounced discrepancies between wards. First, the intervention ‘Collaboration enhancement’ represents 13.90% of all interventions at the urology ward, compared to 9.00% at the neurology ward. Second, ‘Fluid/electrolyte management’ accounts for 5.00% of interventions at the urology ward but represents only 0.90% at the neurology ward. These variations underscore the unique operational characteristics of each ward, which must be considered when studying their work processes.

3.2.2. Overview of the Registrations dataset

The Registrations dataset, containing 393 records, reflects observed registrations by a nurse in the electronic health records. A total of 34 distinct registration types are identified. An overview of the types that were recorded at least five times is provided in the treemap in Figure 3. This cutoff was necessary to ensure the readability of the graph.

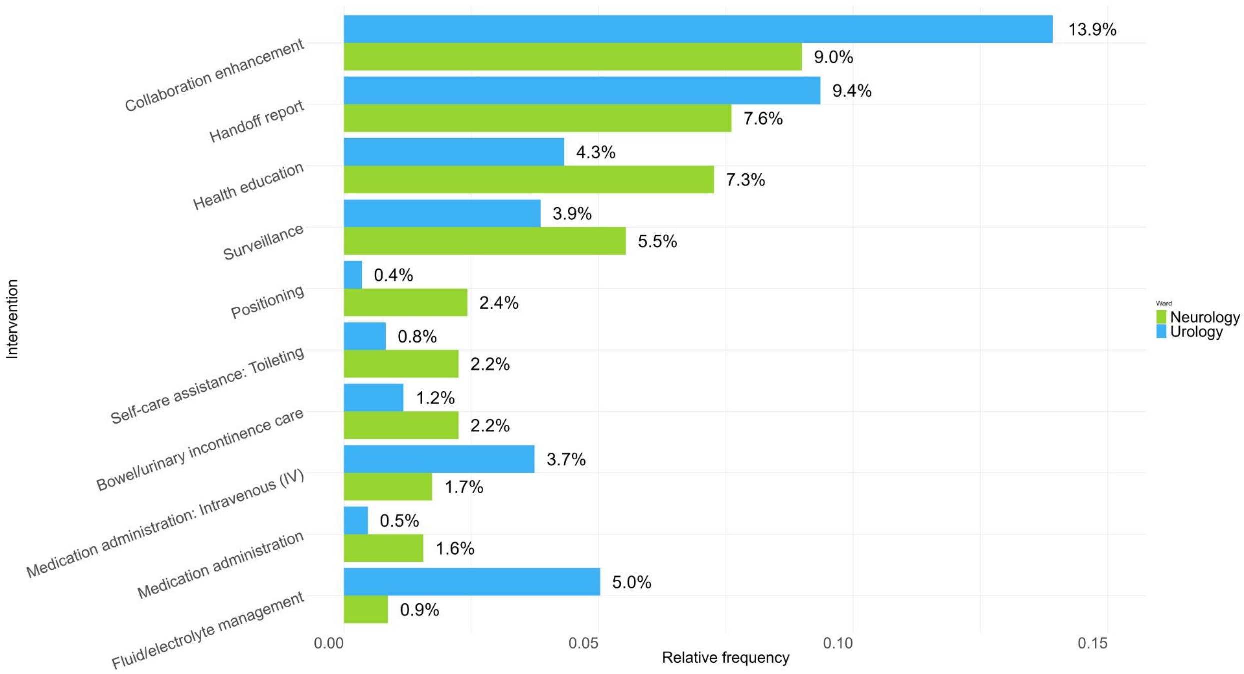

Figure 2: Comparison of intervention frequencies between the Urology and Neurology ward

For the sake of completeness, Appendix C lists all the recorded information types with their frequency. Similar to the observed interventions, particular types of registrations are observed more frequently than others. The ten most frequently recorded types of information account for 94.90% (373 out of 393) of all entries. The 'Administered medication' is the most frequent registration type, representing 126 records (32.06% of all records), followed by the registration of 'Performed activities into the activity plan' (77 records) and entering 'Vital signs values' (49 records). All other registration types are recorded less than 20 times. Comparison between the two observed wards revealed two notable differences. Firstly, registrations related to the activity plan account for 24.80% of entries at the urology ward, compared to only 16.30% at the neurology ward. Additionally, recording patient condition information constitutes a higher proportion of registrations in the neurology ward (6.70% of all records) compared to the urology ward (1.30%).

In 15.52% of the observed registrations, nurses recorded data for multiple types of interventions provided to one patient, a single intervention provided to multiple patients, or multiple interventions for multiple patients. The occurrence of this multi-registration behavior varied depending on the type of data entered into the electronic health records. Notably, the administration of medication was never recorded for multiple patients simultaneously. In contrast, the recording of preparing medication was almost always done for multiple patients at the same time (with only one exception on a total of 14 observed records of preparing medication). Information about patients' activity plans was entered by means of multi-registration in about half (55.26%) of its observed instances. No notable differences were observed in multi-registration behavior between wards.

3.3. (Mis)match between interventions and registrations

Each entry in both datasets carries a timestamp, respectively a timestamp at which an intervention took place in the Interventions dataset and a timestamp at which an electronic health record registration took place in the Registrations dataset. As links between both datasets were captured during data collection, it was possible to link interventions with their registrations and calculate the time gaps between two corresponding occurrences.

3.3.1. Coverage of intervention types by observed registrations

Analyzing which intervention types are covered by observed registrations revealed that 24 intervention types are never linked to an entry in the Registrations dataset. This raised the question if these interventions are simply never

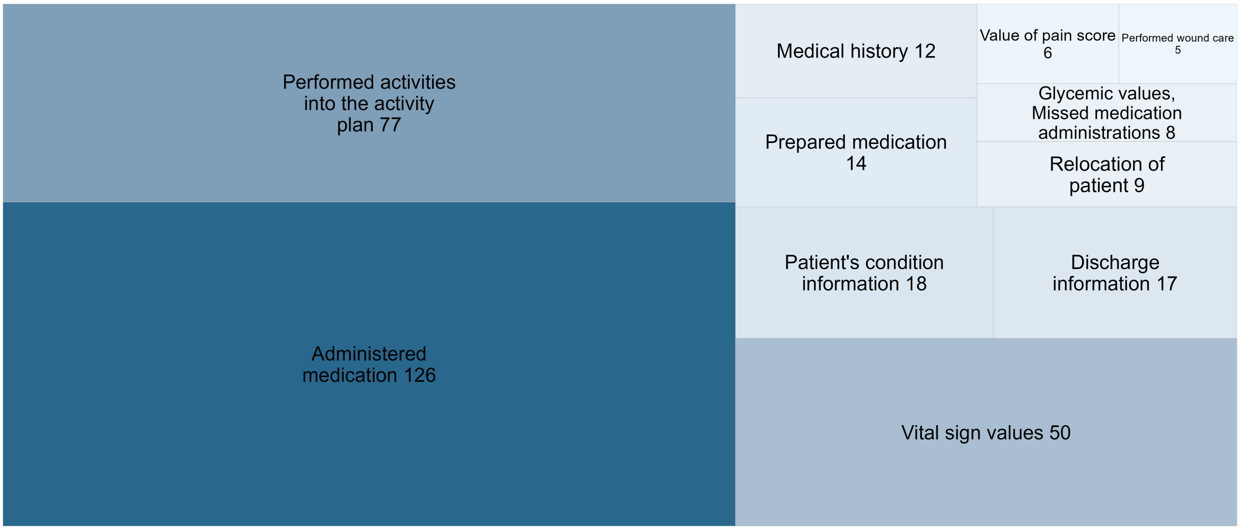

Figure 3: Treemap on the frequency of different registration types observed more than five times

registered by nurses or if their registrations could not be observed. To gain a deeper understanding, an overview of these interventions was provided to a participating nurse for feedback. Appendix D lists all intervention types for which an electronic health record registration is never observed, alongside the nurse's input on whether such interventions are typically registered. According to the nurse, 12 interventions are never registered in the electronic health records among which 'Environmental management: Safety' stood out as it occurred 79 times. One could hypothesize that the registration of the remaining 12 interventions might not be consistently registered in the electronic health records, even though the supporting information system allows the registration of these interventions. However, the available data does not allow us to verify whether this actually holds.

3.3.2. Timing of interventions and registrations

Analyzing the timing of interventions and registrations within shifts reveals differences between shifts. In afternoon shifts, interventions and their registrations were predominantly clustered towards the end of the shift, while interventions were more evenly distributed throughout the night shift. Morning shifts displayed a different pattern, with a high frequency of both interventions and registrations at the start, with a noticeable decrease as the shift progresses.

A comparison of the distribution of interventions and registrations between both wards revealed a difference in the way wards organize their work. At the urology ward, larger volumes of interventions and registrations occurred at three distinct moments after the start of the shift (at approximately 175, 325 and 475 minutes after its start). In contrast, at the neurology ward only two peaks were observed (around 175 and 475 minutes).

Of the 393 observed registrations, 313 (79.64%) could be directly linked to an observed intervention, enabling an analysis of the time gap between the execution of the intervention and its entry into the electronic health records. Summary statistics of the gap between electronic health record registration time and the intervention execution time are provided in Appendix E. If a registration happens after performing an intervention, the median time delay is 1.03 minutes. In contrast, if the registration precedes the intervention, the median time delay between both is only 0.57 minutes.

To reveal whether particular intervention types tend to be registered before or after their execution, Figure 4 depicts the timing of intervention registrations across different Nursing Intervention Classification classes, using the Freedman-Diaconis rule to set the bin-widths to 0.65 minutes (Freedman and Diaconis, 1981). As described in Section 2.3, Nursing Intervention Classification classes are clusters of more specific Nursing Intervention Classification interventions. This choice, as well as the choice to only include classes with over 10 occurrences, was made to ensure the figure's clarity. Figure 4 reveals that interventions classified as 'Drug management' are recorded before, simultaneously and after their execution. It is the only class with pre-execution registrations. Closer inspection of this Nursing Intervention Classification class revealed that interventions regarding medication administration are predominantly

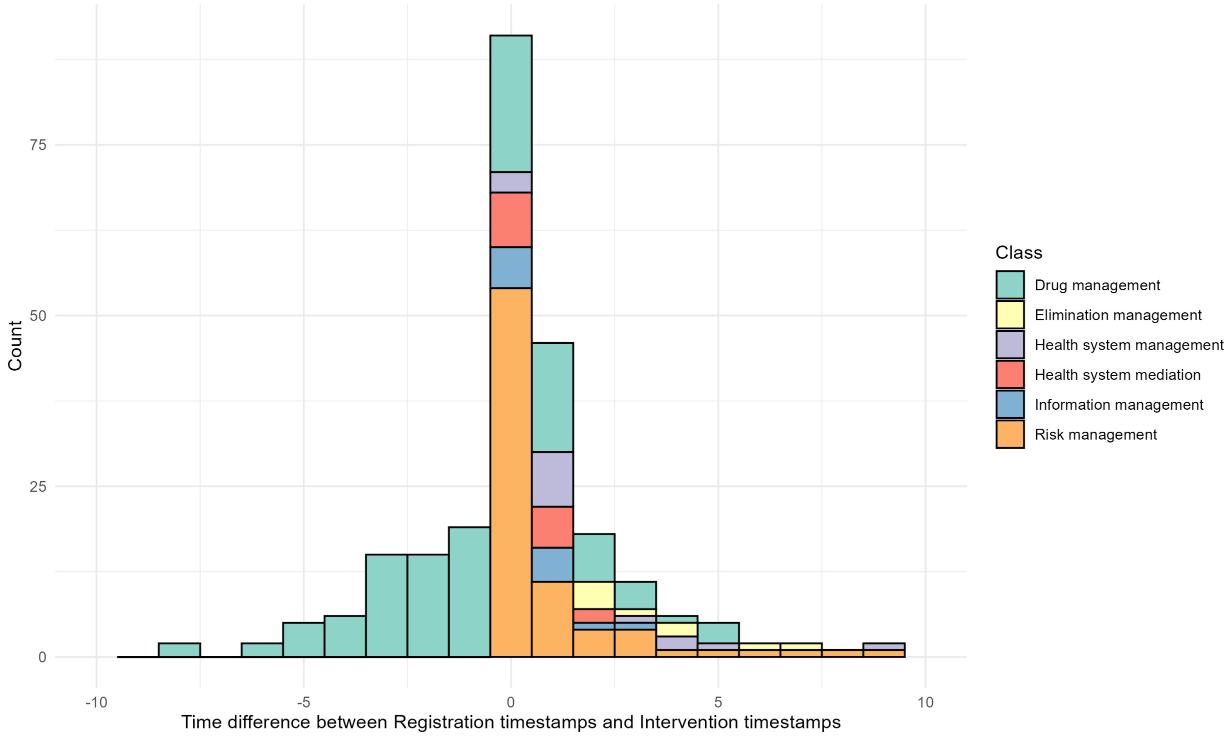

Figure 4: Classes of Nursing Intervention Classification interventions registered before (negative time difference values) or after (positive time values) their execution

registered in advance (73.91% of all medication administration interventions). This is consistent with the insights shared by the nurse during the interview, who explained that medication must first be prepared for patients — an intervention recorded in the electronic health records. Only after this step can the medication be handed to patients, which is documented as a distinct type of entry in the electronic health records. In contrast, 'Risk management' interventions account for most of the registrations occurring simultaneously with the intervention execution. Closer inspection of the distinct interventions captured by this class, revealed that the intervention 'Vital signs monitoring' is almost always registered simultaneously with the intervention execution as the monitoring equipment has a direct link to the electronic health records, thus eliminating the need for manual entry.

A final insight on the link between registrations and their interventions relates to the distribution of the time delay between both. 73 out of 313 (23.32%) observed interventions linked with their registration were labeled as being registered simultaneously with their execution. An intervention was classified as simultaneously registered if its registration occurred within 0.65 minutes of its execution, as indicated by the bin clustered around 0 in Figure 4. The bulk of interventions (204 out of 313 or 65.18%) were registered within a 10-minute interval before or after their execution. The remaining 36 interventions (11.50%) were registered more than 10 minutes before or after the intervention.

As only 313 out of 393 registrations could be linked to an intervention, the remaining 80 records were examined more closely. Closer inspection of these records, together with the insights retrieved from the in-depth discussion with the nurse, surfaced that 38 out of these 80 records involved nurses registering information about interventions performed by a colleague of whom interventions were not observed during that shift. Another 32 records involved multi-registration behavior of entering information into the patients' activity plans, during which the observer could not link specific registrations to corresponding interventions. For the final 10 registrations that were not linked to an intervention, the observer was just not able to determine the specific intervention to which the registration belonged.

As highlighted above, 32 records in the Registrations dataset involved the multi-registration behavior in activity plans (i.e. the individualized care plan specifying the nursing interventions to be performed for a particular patient).

Each of these records relates to multiple interventions in the Interventions dataset. To investigate the (mis)match between the time an intervention took place and its electronic health record registration on a larger number of data points, a heuristic is applied within each observed shift to establish additional connections between interventions and registrations. For every intervention in the Interventions dataset that is part of the activity plan, without a corresponding entry in the Registrations dataset, the following heuristic is applied: connect this intervention to the registration marked as 'Activity plan' with the timestamp that is nearest to the intervention's timestamp. Applying this heuristic constitutes the most optimistic situation as interventions are linked to the potential registration time that is closest in time. Using this heuristic, an additional 240 interventions could be linked to a registration. Consequently, an imputed dataset was created that contained 553 interventions linked to an electronic health record registration.

Analysis of the imputed dataset revealed that 255 out of 553 (or 46.11%) registrations occurred more than 10 minutes before or after the intervention. 77 out of 553 (or 13.92%) interventions were registered simultaneously. The final 221 registrations (or 39.96%) occurred within 10 minutes of the corresponding intervention. This finding suggests a greater deviation between the intervention time and the registration time than established based on observational data alone.

3.3.3. Relationship between interventions and registrations

The relationship between the type of registration in the electronic health records and the corresponding intervention is not always unique. Some types of registrations correspond to multiple interventions and a specific intervention can lead to different types of data being entered. Figure 5, Pane B, shows all interventions registered under two or more registration types. For example, the intervention 'Urinary retention care' can be documented in the patient's activity plan or as the residual volume of urine left in the bladder recorded as a vital sign. Similarly, Figure 5, Pane A, presents electronic health record registration types that correspond to multiple interventions. For instance, the registration type 'Vital sign values' might represent either the intervention of the preparation of a handoff report or the actual measurement of those vital signs. For some interventions, the used registration type depends on the nurses' habit regarding the use of the electronic health records. These findings underscore the complex relationship between nursing interventions and the data entered in the hospital information system.

The specific challenges encountered will depend on factors such as the level of granularity required to define interventions for the process mining use case. For example, a coarse level of granularity may group distinct interventions under a single registration type, potentially obscuring variations in workflows. In this use case, the registration type 'Administered medication' illustrates this challenge. This registration type could encompass both the administration of intravenous medication and oral medication, two interventions that differ in required skill set and the time needed for completion. Such aggregation can hinder the nuanced understanding of workflows of nurses.

Conversely, a coarser level of granularity might aggregate distinct interventions under a single data type, potentially masking meaningful variations in workflows. The choice of granularity thus significantly impacts both the interpretability of the findings and the feasibility of deriving actionable insights from the process mining analysis.

4. Discussion

To obtain accurate insights into healthcare processes from electronic health record data, it is important that electronic health record registrations adequately match what happens in reality. This observational study at two wards in a Belgian hospital focuses specifically on better understanding the (mis)match between nursing interventions and observed electronic health record registrations.

4.1. Key findings

The analysis of the observational data highlights three key mismatches between electronic health record registrations and actual nursing work and enables a comparison between two distinct hospital wards. Firstly, *some nursing interventions will not be reflected in the electronic health records*. For example, interventions regarding the nutrition management of patients or interventions related to meetings are not recorded in the electronic health records. As a result, these interventions will be missing from datasets derived solely from the electronic health records, potentially leading to incomplete or biased insights when analyzing care processes. Among the nursing interventions that are not intended to be documented in the electronic health records, two stand out: 'Environmental Management: Safety' and 'Teaching: Individual/Staff Development'. Remarkably, these interventions rank among the top ten most frequently performed interventions by nurses. The relevance of this finding becomes even more apparent when taking into account

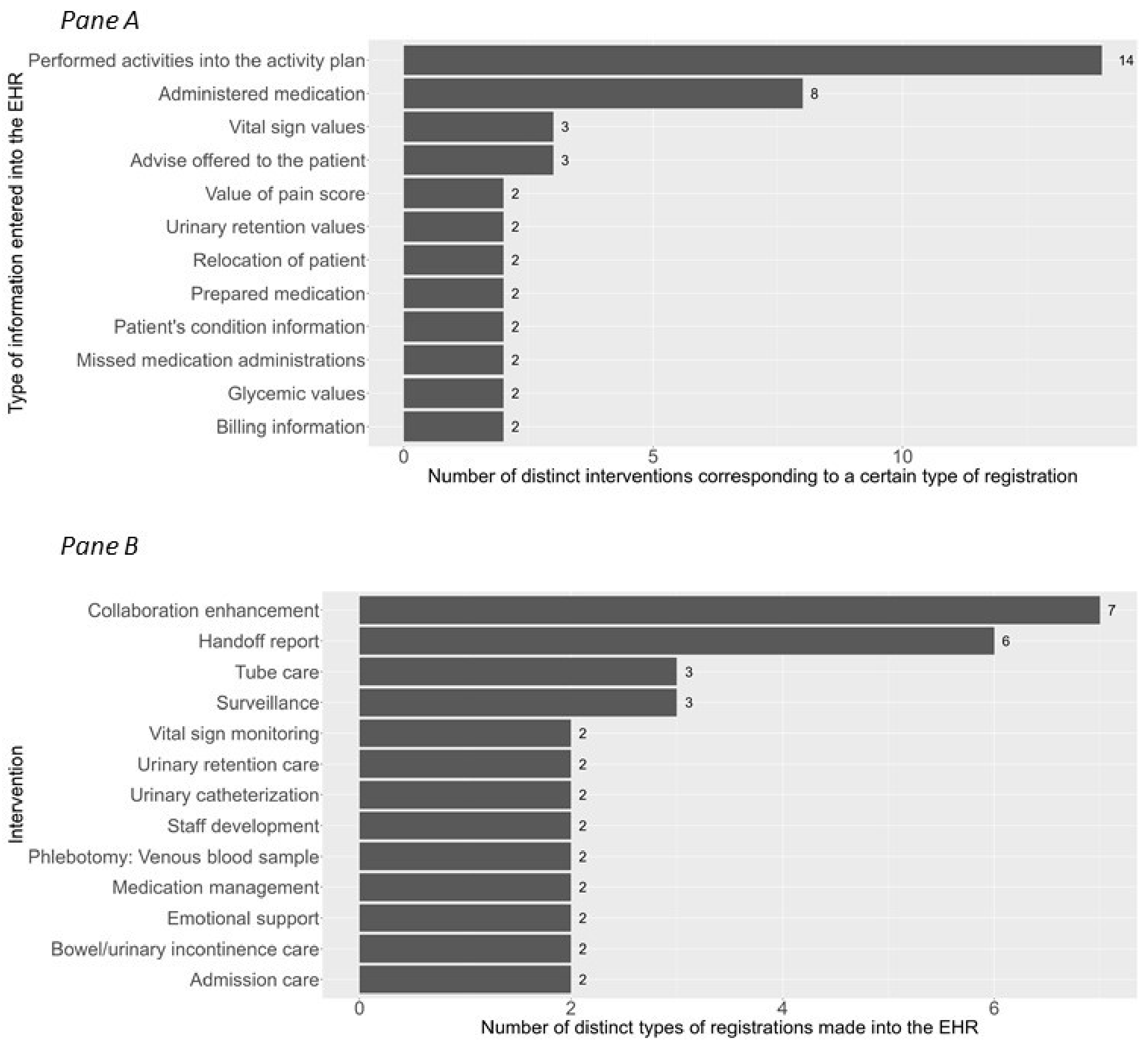

Figure 5: Interventions that were registered as multiple distinct registration types in the hospital information system (Pane A) and types of registrations that could relate to multiple performed interventions (Pane B)

the overall distribution of observed interventions: the ten most frequently observed interventions account for 60.60% of all recorded nursing interventions. This suggests that a relatively small set of interventions dominates daily nursing practice, making it even more concerning that some of these high-frequency interventions are not captured in the electronic health records. It reveals a considerable blind spot when relying solely on electronic health record data to study nursing work, as important aspects of nursing practice remain invisible in such datasets. These findings align with the results obtained by Fore et al. (2019) as their observational study also showed that many nursing interventions were not captured in the electronic health records, highlighting the disconnect between actual nursing practice and what is represented in electronic health record data.

Secondly, *only a limited number of interventions are registered in the electronic health records at the time they are actually performed*. Only 23.32% of all 313 interventions that could be linked to their registration were registered simultaneously with their execution. A further 65.78% were registered within a 10-minute window before or after their execution. As a result, the remaining 10.90% of interventions were registered more than 10 minutes apart from their execution. To refine the analysis, a heuristic was developed enabling the linkage of an additional 240 interventions with their respective electronic health record entries. Analysis of this imputed dataset reveals that only 46.11% of registrations occur within the 10-minute window, suggesting an even larger deviation in time between interventions and their registrations. Accurate information on the timing of nursing interventions is essential for gaining meaningful insights into how nursing care is delivered. This study revealed a mismatch between the actual moment an intervention is performed and the time it is registered in the electronic health records. Such discrepancies complicate efforts to analyze nursing workflows based on electronic health record data, particularly when examining time-sensitive aspects such as potential delays, workload distribution, or care efficiency. As a result, relying solely on electronic health record registrations to study nursing practice may lead to an incomplete or distorted understanding of real-world care delivery.

Thirdly, *the relationship between interventions and electronic health record registrations is often complex*. In this study, the observational data from the Interventions and Registrations dataset did not demonstrate a one-to-one relationship. A single intervention type could be recorded as various types of registrations in the electronic health records. Moreover, a single type of registration in the electronic health records could correspond to multiple interventions. As a result, it becomes challenging to determine how nursing work is reflected in the electronic health records, particularly when attempting to reconstruct care processes. For some interventions, this can even depend on nurses' habits regarding the use of the electronic health records as they might register the same intervention in different ways. These findings underscore the importance of understanding how data are entered into electronic health record systems. Observational research plays a critical role in uncovering how care delivery is translated into digital records. This knowledge is essential not only for data analysts working with electronic health record data, but also for healthcare leaders and policymakers who rely on such data to support decision-making and quality improvement.

Finally, *the observational study took place at two distinct wards, allowing their comparison*. No notable differences were observed between the wards regarding the time gap between electronic health record registrations and the intervention executions, suggesting that the observed patterns might hold more broadly within the healthcare domain. However, differences were found in the type of interventions performed on each ward. For instance, fluid and electrolyte management interventions accounted for 5.00% of all interventions in the urology ward, whereas these interventions only cover 0.90% of the interventions in the neurology ward. The presence of such differences was expected given the different patient populations served by each ward. Additionally, differences in work organization were observed between wards. At the urology ward, more interventions and registrations occured at three distinct moments after the start of a shift, whereas the neurology ward only shows two such distinct periods. These findings emphasize the importance of considering the specific healthcare context when conducting process mining studies in healthcare settings.

4.2. Value of observational research for understanding electronic health record data

This study underscores the importance of going beyond the data itself to truly understand what data represent. While the importance of data quality assessment is well established within the process mining literature (Martin, 2021; Martin, Fischer, Kerpedzhiev, Goel, Leemans, Röglinger, van der Aalst, Dumas, La Rosa and Wynn, 2021; Munoz-Gama et al., 2022), our findings suggest that such assessments should go beyond the data itself to consider how the data are generated in practice. Instead of solely looking at the data to assess its quality, it is crucial to consider how work processes are carried out in practice and to what extent the supporting systems capture these processes without leaving important blind spots. Observing process participants, such as nurses, can provide valuable insights into the origins of the data. For example, some nurses register medication preparation immediately after completing the task, while others prefer to prepare all medications first and record the information later in the electronic health records. These variations in documentation behavior can introduce timestamp inaccuracies, which in turn affect the reliability of time-sensitive process analyses. Therefore, understanding the context in which data are recorded is crucial for correctly interpreting and using electronic health record data in process mining efforts.

A better understanding of the actual work environment and documentation practices can help to anticipate and mitigate data quality issues. The presented use case highlights that timestamps recorded in the electronic health records often do not accurately reflect the timing of nurses' interventions. Identifying this data quality issue can inspire solutions to mitigate its impact. For instance, hospitals could adopt a bedside registration approach or invest in equipment integrated with the electronic health records to automatically record results. Another significant data quality issue identified in this study is the incompleteness of electronic health records due to interventions that are not captured. The identification of these interventions could inspire innovative strategies to ensure their registration. For example, leveraging other data sources beyond the electronic health records, such as telephone records to visualize interdepartmental collaboration, could provide valuable complementary insights into nursing work that remain invisible in current datasets.

Ultimately, the presented results are valuable not only for data analysts or process improvement teams, but also for healthcare leaders and policymakers. Without insight into how data are generated, there is a risk that electronic health record-based dashboards, reports, or analyses are interpreted without a full understanding of the context in which the recorded work takes place. Such interpretations are at risk of being incomplete or misleading. Observational research helps bridge this gap: it fosters mutual understanding between those who work with data and those who produce it.

4.3. Limitations

The obtained results need to be reflected against the study's limitations. Firstly, all collected data originates from extensive, yet finite observations. Hence, we cannot guarantee the completeness of the collected data. However, extensive observations have been conducted until saturation has been reached, i.e. until no significant new information was collected during an observed shift. This reduces the likelihood of missing important information. Secondly, all data were collected at a single hospital, which may limit the generalizability of the findings. However, observations were performed across two distinct wards, each with its own specific patient population and separate staff. The comparison between the two wards provides insights into potential differences or similarities within healthcare organizations. Nevertheless, further research is needed to confirm these findings and to better understand their generalizability across different healthcare settings. Thirdly, a part of the records in the Registrations dataset (20.36%) could not be linked to their corresponding record in the Interventions dataset due to multi-registration behavior, reducing the number of records for which an exact time gap between intervention execution and electronic health record registration could be calculated. However, the developed heuristic enabled establishing additional connections between registration times and interventions, albeit representing the most optimistic scenario (with interventions being registered in the electronic health records as close as possible to their execution). Finally, as with any direct observation, the Hawthorne effect might caused nurses to change their behavior because they knew they were being observed (Lopetegui, Yen, Lai, Jeffries, Embi and Payne, 2014). All participating nurses were thoroughly informed in advance about the goals of the study and, e.g., assured that no personal information would be recorded, reducing the risk that this effect would materialize. Moreover, the regular presence of the observer at the ward could have made the nursing team feel accustomed to the study situation, increasing the likelihood that nurses exhibit normal working behavior.

5. Conclusion

This study examined the (mis)match between nursing interventions and their registration in the electronic health records by means of direct observations. The findings show that not all interventions are registered, that time gaps between execution and documentation are common, and that the link between nursing activities and their documentation is often inconsistent, where the same intervention can be registered in different ways, and one type of registration may represent more than one type of intervention. These mismatches raise important concerns about the completeness and accuracy of electronic health record data as a basis for understanding how nurses get work done in a work process.

Observational research proved valuable in identifying these limitations and offered essential insight into how nursing work is carried out and documented in real-world settings. Such understanding is important not only for researchers using electronic health record data, but also for healthcare leaders and policymakers who rely on these data to inform decisions around care quality, staffing and work process improvement.

Future research could explore how documentation practices vary across different care contexts and how data quality might be improved without increasing nurses' administrative burden. Possible strategies include bedside documentation, partial automation or the integration of complementary data sources (e.g. telephone records). Embedding observational methods into data-driven research and data quality improvement efforts can help ensure that digital records provide a more accurate and meaningful reflection of nursing work.

6. Conflicts of interest

Conflicts of interest: none.

A. Schedule of observed nursing shifts across different wards (Neurology and Urology) and days.

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Morning shift	11 March 2024 18 March 2024			7 March 2024 21 March 2024		16 March 2024 30 March 2024	
Afternoon shift		26 March 2024	13 March 2024 10 April 2024		23 February 2024 1 March 2024		3 March 2024 24 March 2024
Night shift					8 March 2024		

B. Observed nursing interventions and their frequency

Nursing intervention	n
Collaboration enhancement	171
Handoff report	124
Documentation	104
Environmental management: safety	79
Health education	79
Medication administration: oral	70
Surveillance	65
Supply chain management	62
Vital sign monitoring	58
Teaching: individual/staff development	57
Medication management	54
Fluid/electrolyte management	48
Medication administration: Intravenous (IV)	42
Self-care assistance: Feeding	40
Staff development	37
Environmental management: comfort	35
Admission care	24
Bowel/urinary incontinence care	23
Self-care assistance: Toileting	20
Environmental management: Worker safety	18
Positioning	17
Transport: Intrafacility	17
Phlebotomy: Venous blood sample	15
Bathing	14
Discharge planning	14
Wound care	14
Health care information exchange	13
Medication administration	13
Neurologic monitoring	11
Laboratory data interpretation	7
Urinary catheterization	7
Medication administration: subcutaneous	6
Self-care assistance: IADL	6
Heat/cold application	5
Tube care	5
Urinary retention care	5
Emotional support	4
Examination assistance	4
Health system guidance	4
Oral health maintenance	4
Staff supervision	4
Bladder irrigation	3
Dressing	3
Hyperglycemia management	3
Intravenous (IV) therapy	3
Oxygen therapy	3
Research data collection	3

Continuation of Table B

Nursing intervention	n
Behavior modification	2
Cardiac care	2
Hair and scalp care	2
Medication administration: Skin	2
Documentation: meetings	1
Embolus precautions	1
Nausea management	1
Nutrition management	1
Physical restraint	1
Preceptor: Student	1
Seizure management	1
Specimen management	1

C. Types of information entered in the electronic health records and their frequency

Information entered in the electronic health records	n
Administered medication	126
Performed activities into the activity plan	77
Vital sign values	49
Patient's condition information	18
Discharge information	17
Medication prepared	14
Medical history	12
Relocation of patient	9
Missed medication administrations	8
Glycemic values	8
Value of pain score	6
Performed wound care	5
Results of the NIHSS	4
Urinary retention values	3
Specimen information	3
Offered nutrition support	3
Billing information	3
Admission information	3
Specifics about administered oxygen therapy	2
Results from pre-examination questionnaires	2
ECG results	2
Changes in medication regimen	2
Advise offered to the patient	2
Tube care	1
Temperature values	1
Results of the neurological observation	1
Restock of incontinence material	1
Performed tube care	1
Offered patient education	1
Notes of patient absence	1
Key concerns for patients	1
IV-line caretaking	1
Fixation specifics	1
Examination findings	1
Completion of the EHR	1
Changes in the activity plan	1
Appointments made for patients	1

D. Nurse feedback on electronic health record registration of observed intervention types for which a registration is never observed

Observed intervention types for which a registration is never observed	electronic health record registration according to nurse feedback
Teaching: individual/staff development; Heat/cold application; Environmental management: Safety; Feeding; Environmental management: Worker safety; Nutrition management; Healthcare information exchange; Research data collection; Documentation: Meetings; Preceptor: Student; Staff supervision; Self-care assistance: Bathing/hygiene	No
Supply chain management	Partially - only medication ordering
Self-care assistance: IADL; Embolus precautions; Behavior modification; Intravenous (IV) therapy; Specimen management; Oral health maintenance; Health system guidance; Hair and scalp care; Dressing; Nausea management; Seizure management	Yes

E. Summary statistics of the time difference between the electronic health record registration and the intervention execution (expressed in minutes).

	Average	Median	Minimum	Maximum	Q1	Q3
Registration after execution	9.83	1.03	0	297.99	0.13	4.46
Registration before execution	2.48	0.57	0	55.25	0	2.44

References

van der Aalst, W.M.P., 2016. Process Mining: Data Science in Action. Springer, Heidelberg.

van der Aalst, W.M.P., 2022. Process mining: a 360 degree overview, in: van der Aalst, W.M.P., Carmona, J. (Eds.), Process Mining Handbook. Springer, Cham. volume 448 of *Lecture Notes in Business Information Processing*, pp. 3–34.

Adler-Milstein, J., DesRoches, C.M., Kralovec, P., Foster, G., Worzala, C., Charles, D., Searcy, T., Jha, A.K., 2015. Electronic health record adoption in us hospitals: progress continues, but challenges persist. Health Affairs 34, 2174–2180.

Alghamdi, M., 2016. Nursing workload: a concept analysis. Journal of Nursing Management 24, 449–457.

Baker, K., Dunwoodie, E., Jones, R.G., Newsham, A., Johnson, O., Price, C.P., Wolstenholme, J., Leal, J., McGinley, P., Twelves, C., et al., 2017. Process mining routinely collected electronic health records to define real-life clinical pathways during chemotherapy. International journal of medical informatics 103, 32–41.

Benevento, E., Dixit, P.M., Sani, M.F., Aloini, D., van der Aalst, W.M., 2019. Evaluating the effectiveness of interactive process discovery in healthcare: a case study. Lecture Notes in Business Information Processing 362, 508–519.

Bose, R.J.C., Mans, R.S., Van Der Aalst, W.M., 2013. Wanna improve process mining results?, in: 2013 IEEE symposium on computational intelligence and data mining (CIDM), IEEE. pp. 127–134.

Butcher, H., Bulechek, G., Dochterman, J., Wagner, C., 2018. Nursing Interventions Classification 7e (NIC). Elsevier, Missouri.

Cho, M., Song, M., Park, J., Yeom, S.R., Wang, I.J., Choi, B.K., 2020. Process mining-supported emergency room process performance indicators. International Journal of Environmental Research and Public Health 17, 6290.

Cremerius, J., König, M., Warmuth, C., Weske, M., 2021. Patient discharge classification based on the hospital treatment process, in: International Conference on Process Mining, Springer International Publishing Cham. pp. 314–326.

De Roock, E., Martin, N., 2022. Process mining in healthcare –an updated perspective on the state of the art. Journal of Biomedical Informatics 127, 103995.

Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A., et al., 2013. Fundamentals of business process management. volume 1. Springer.

Fennelly, O., Grogan, L., Reed, A., Hardiker, N.R., 2021. Use of standardized terminologies in clinical practice: A scoping review. International Journal of Medical Informatics 149, 104431.

Fore, A., Islim, F., Shever, L., 2019. Data collected by the electronic health record is insufficient for estimating nursing costs: An observational study on acute care inpatient nursing units. International Journal of Nursing Studies 91, 101–107.

Fox, F., Aggarwal, V.R., Whelton, H., Johnson, O., 2018. A data quality framework for process mining of electronic health record data, in: Proceedings of the 2018 IEEE International Conference on Healthcare Informatics, IEEE. pp. 12–21.

Fox, F., Whelton, H., Johnson, O.A., Aggarwal, V., 2023. Dental extractions under general anesthesia: new insights from process mining. JDR Clinical & Translational Research 8, 267–275.

Freedman, D., Diaconis, P., 1981. On the histogram as a density estimator: L 2 theory. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 57, 453–476.

Häyrinen, K., Saranto, K., Nykänen, P., 2008. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. International journal of medical informatics 77, 291–304.

Van den Heede, K., Balcaen, K., Bouckaert, N., Bruyneel, L., Cornelis, J., Sermeus, W., Van de Voorde, C., 2023. Improving hospital nurse staffing during the pandemic: Implementation of the 2019 fund for health care staff in belgium. Health policy 128, 69–74.

Huang, Z., Dong, W., Ji, L., Yin, L., Duan, H., 2015. On local anomaly detection and analysis for clinical pathways. Artificial Intelligence in Medicine 65, 167–177.

Jianxun, C., Arkorful, V.E., Shuliang, Z., 2021. Electronic health records adoption: Do institutional pressures and organizational culture matter? Technology in Society 65, 101531.

Jonk, J., Schaller, M., Netzer, M., Pfeifer, B., Ammenwerth, E., Hackl, W., . Process mining of nursing routine data: Cool, but also useful? 293, 137–144.

Lopetegui, M., Yen, P.Y., Lai, A., Jeffries, J., Embi, P., Payne, P., 2014. Time motion studies in healthcare: what are we talking about? Journal of biomedical informatics 49, 292–299.

Mann, S., 2013. Research methods for business: A skill-building approach. Leadership & Organization Development Journal 34, 129–145.

Mans, R.S., van der Aalst, W.M.P., Vanwersch, R.J.B., 2015. Process mining in healthcare: evaluating and exploiting operational healthcare processes. Springer, Cham.

Martin, N., 2021. Data quality in process mining. Interactive process mining in healthcare , 53–79.

Martin, N., De Weerdt, J., Fernández-Llatas, C., Gal, A., Gatta, R., Ibáñez, G., Johnson, O., Mannhardt, F., Marco-Ruiz, L., Mertens, S., et al., 2020. Recommendations for enhancing the usability and understandability of process mining in healthcare. Artificial Intelligence in Medicine 109, 101962.

Martin, N., Fischer, D.A., Kerpedzhiev, G.D., Goel, K., Leemans, S.J.J., Röglinger, M., van der Aalst, W.M.P., Dumas, M., La Rosa, M., Wynn, M.T., 2021. Opportunities and challenges for process mining in organizations: results of a Delphi study. Business & Information Systems Engineering 63, 511–527.

Munoz-Gama, J., Martin, N., Fernandez-Llatas, C., Johnson, O.A., Sepúlveda, M., Helm, E., Galvez-Yanjari, V., Rojas, E., Martinez-Millana, A., Aloini, D., et al., 2022. Process mining for healthcare: characteristics and challenges. Journal of Biomedical Informatics 127, 103994.

Noshad, M., Rose, C.C., Chen, J.H., 2022. Signal from the noise: a mixed graphical and quantitative process mining approach to evaluate care pathways applied to emergency stroke care. Journal of biomedical informatics 127, 104004.

Partington, A., Wynn, M., Suriadi, S., Ouyang, C., Karnon, J., 2015. Process mining for clinical processes: a comparative analysis of four australian hospitals. ACM Transactions on Management Information Systems 5, 1–18.

Reinkemeyer, L., 2020. Process Mining in Action: Principles, Use Cases and Outlook. Springer, Heidelberg.

Rinner, C., Helm, E., Dunkl, R., Kittler, H., Rinderle-Ma, S., 2018. An application of process mining in the context of melanoma surveillance using time boxing. Lecture Notes in Business Information Processing 342, 175–186.

Rojas, E., Munoz-Gama, J., Sepúlveda, M., Capurro, D., 2016. Process mining in healthcare: A literature review. Journal of Biomedical Informatics 61, 224–236.

Stalpers, D., Schoonhoven, L., Dall'Ora, C., Ball, J., Griffiths, P., 2025. 'entanglement of nursing care': A theoretical proposition to understand the complexity of nursing work and division of labour. International Journal of Nursing Studies , 104995.

Stefanini, A., Aloini, D., Benevento, E., Dulmin, R., Mininno, V., 2018. Performance analysis in emergency departments: a data-driven approach. Measuring Business Excellence 22, 130–145.

Suriadi, S., Andrews, R., ter Hofstede, A.H., Wynn, M.T., 2017. Event log imperfection patterns for process mining: towards a systematic approach to cleaning event logs. Information Systems 64, 132–150.

Vanbrabant, L., Martin, N., Ramaekers, K., Braekers, K., 2019. Quality of input data in emergency department simulations: Framework and assessment techniques. Simulation Modelling Practice and Theory 91, 83–101.

Yoo, S., Cho, M., Kim, E., Kim, S., Sim, Y., Yoo, D., Hwang, H., Song, M., 2016. Assessment of hospital processes using a process mining technique: Outpatient process analysis at a tertiary hospital. International journal of medical informatics 88, 34–43.

Zerbino, P., Stefanini, A., Aloini, D., 2021. Process science in action: A literature review on process mining in business management. Technological Forecasting and Social Change 172, 121021.