dHealth 2022
G. Schreier et al. (Eds.)

doi:10.3233/SHTI220360

Process Mining of Nursing Routine Data: Cool, but also Useful?

Julian JONK[a,b,1], Michael SCHALLER[a], Michael NETZER[a], Bernhard PFEIFER[a], Elske AMMENWERTH[a] and Werner HACKL[a]

[a] *UMIT TIROL – Private University for Health Sciences, Medical Informatics and Technology, Institute of Medical Informatics, Hall in Tirol, Austria*
[b] *Amsterdam UMC, Location AMC, Department of Medical Informatics, Amsterdam, The Netherlands*

Abstract. Background: Process mining is a promising field of data analytics that is yet to be applied broadly in healthcare. It can streamline the care process, leading to a higher quality of care, increased patient safety and lower costs. Objectives: To get deeper insights into the emergence and detectability of delirium in a gerontopsychiatric setting. Methods: We use process mining to create process models from routinely collected, anonymised nursing data from two gerontopsychiatric wards. We analyse these models to get a longitudinal view of the care processes. Results: The process models comprise all activities during patients' stays but are too extensive and challenging to interpret due to the wide variation in care paths. Although the models give insight into frequent paths and activities, they are insufficient to explain the emergence of delirium meaningfully. No apparent difference between stays with or without delirium could be detected. Conclusion: Conducting process mining on routinely collected data is easy, but the interpretation of the results was a challenge. We identified four limitations associated with using this data and gave recommendations on adapting it for further analysis.

Keywords. Data Mining, Delirium, Process Assessment, Delivery of Health Care

1. Introduction

1.1. Re-use of Medical Data

Medical data collection and registration have shaped healthcare as we know it and significantly increased the quality of care. Since the start of digital data collection, medical data has grown exponentially, with approximately 30% of the world's data volume generated in healthcare facilities [1]. Electronic Health Record (EHR) adaption significantly encourages this trend. Already, physicians spend 49% of their day on administration, working in the EHR, and desk work [2]. Furthermore, administrative expenses account for approximately 15% to 25% of national US healthcare expenditure [3]. All in all, much time, effort, and money go into maintaining a high-quality EHR, setting the foundation for future research.

In most cases, the data in the EHR comprises highly detailed and structured patient information, mapping the complete patient stay. These characteristics offer the remarkable potential to research and re-use this data, laying the foundations for a higher

[1] Corresponding Author: Julian Jonk, Amsterdam UMC, Amsterdam, The Netherlands, E-mail: j.s.jonk@amsterdamumc.nl

quality of care, better care management, lower healthcare costs, and improved clinical research. The latter is pre-eminent in the medical world, with most healthcare facilities having extensive provisions for researchers to request data for their research. However, most medical research focuses on the outcomes and effectiveness of care rather than the process itself. After all the resources spent to generate this (process) data, few healthcare facilities even use any data-driven analytics to re-use this, mainly because current data analytics techniques are hard to apply to heterogeneous hospital data [4].

One of the techniques that healthcare facilities could apply to give this data a secondary purpose is process mining. Process mining is a relatively young research field that focuses on extracting knowledge from data generated during normal processes [5]. It uses the event logs in the data to build a process model, an overview of all the steps in a process. Process mining is already widely adopted in business, such as resource and time optimisation. Nonetheless, process mining is still in its infancy in healthcare and is primarily used in case studies. When applied correctly, process mining can uncover numerous insights for enhancing the healthcare process, from saving costs to improving patient outcomes [4,6].

1.2. Case study: Delirium Care and Detection

This work will investigate nursing data from a delirium detection project on two gerontopsychiatric wards. Delirium is an acute neuropsychiatric syndrome causing physical, cognitive, and psychiatric abnormalities, observed in approximately one in six hospitalised elderly patients [7]. On top of that, it is a very volatile condition, which makes it hard to detect for both physicians and nurses. Many instruments to assess the risk of delirium exist, but we will focus on the Delirium Observation Screening Scale (DOS) and Confusion Assessment Method (CAM). The DOS is a 13-point dichotomous questionnaire conducted by nurses during patient observation [8]. The CAM is a 9-point questionnaire completed by clinicians during a patient interview [9].

Although both instruments have proven their worth in multiple settings in the past, problems arise when applying them on a gerontopsychiatric ward [10,11]. It is not clear whether these instruments are sensitive and specific enough for this type of patient because of their multimorbidity. Hospitalised elderly have a far greater likelihood of suffering from dementia, acute confusion, and depression than their younger counterparts [12,13]. The symptoms associated with these conditions overlap partly with the symptoms of delirium, which could cause false positives in the screening instruments.

To explore and investigate this, we examined two gerontopsychiatric wards and applied both the DOS and CAM regularly, once at the admission and once every week of the patient’s stay. In a previous study, we compared the outcomes and scores from the screening instruments with the gold standard; the diagnosis of delirium according to the ICD-10 criteria, which was set simultaneously to CAM and DOS. These findings corroborated the presumption mentioned above. In particular, the DOS instrument was very unspecific, resulting in false positives in over 35% of the observations. The CAM outperformed the DOS in terms of accuracy but still did worse compared to the results in the literature [11]. Unfortunately, no clear patterns could be identified in the development of the scores and diagnosis [14]. We found visual analytics techniques to probably be helpful to gain a deeper insight into the different patient stays [15].

Thus, the objective of this work was to apply process visualisation and process mining to get a horizontal perspective on the timeline of the patient stays. We wanted to learn if these automatically built process models can explain the emergence of delirium during patient stays and if we can see differences in the treatment processes for patients with or without delirium.

2. Methods

As a dataset, we used retrospectively collected data from two gerontopsychiatric wards as described in [14] (Ethical approval granted by the Ethical Committee of the Medical University of Innsbruck - EK Nr: 1032/2019). The dataset comprised the DOS and CAM data and data on the ICD-10 diagnoses of delirium as the gold standard. Additionally, data from the electronic documentation of the nursing process, including data from nursing diagnoses (NANDA-I) and nursing interventions (standardised house catalogue).

All data was fully anonymised. No personal information on the patients was included. Even timestamps were removed, and only relative time measures compared to the initial admission timepoint of each stay were included.

The dataset comprised data from all 708 stays between February 2019 and April 2020 on two wards with 185,496 single data items.

2.1. Building the Process Model

We used Fluxicon's Disco (FD) process mining tool to analyse our dataset [16]. We selected FD for this study because of its high ease of use and adaptability for healthcare data. It also has one of the fastest process mining algorithms available, which is useful when applying process mining to complex and diverse medical data.

Because the scores of the detection instruments must be compared to the clinical diagnosis of delirium, we omitted all cases without a delirium diagnosis. We made the data suitable for analysis by labelling all nursing measures, the DOS and CAM scores, and the diagnoses as 'activities' in the process model. In FD, activities are seen as unique steps in the process, whilst events represent each encounter of an activity. The anonymised and random patient identifiers were labelled as 'ID', and the relative timestamps as 'time'. By using relative timestamps, we ensured the data was not retraceable without losing information on the duration of the care process.

The DOS and CAM were conducted upon admission and then weekly as long as the stay lasted. The same was done for the clinical diagnosis. Because each question in the DOS answered with 'yes' adds a point, these scores vary from 0 to 13. We used a cut-off value of three or higher as a positive indication for delirium per earlier research. For the CAM, three different outcomes for delirium occur in the data: yes, no, and possibly.

We first examined the dataset (.csv file) to check for conformity and completeness. The file was then loaded into FD, with all cases and actions successfully being converted into the process model. FD has two sliders to set a threshold on the percentages of activities and paths shown in the process model; we tested and put these to an optimal value for completeness and comprehensibility. We applied the animation in FD to get a real-time view of how the patients flow through the process. We used the 'statistics' function to examine the case duration and the frequency of activities. Finally, we used the 'cases' function to overview the care process of one or identical case(s) if this was deemed relevant by observing the process model.

We made three visualisations of the process model: one from the complete dataset and two with subsets of the data. The first subset contained all patients that received a clinical diagnosis of delirium during their stay. The second subset of patients continuously received a positive DOS throughout their stay without developing delirium. These are essentially the patients that receive a false-positive indication their whole visit. The subsets were constructed in the FD platform itself by in- and excluding activities.

3. Results

3.1. Data Characteristics

The dataset contained 185,496 events attributable to 456 distinct activities during the patient stays. 516 patients received only negative delirium diagnoses (no delirium according to ICD-10 criteria) during their stay. 37 patients received at least one positive diagnosis of delirium. 155 patients did not receive a clinical diagnosis for delirium at all. The number of stays included in this research was, therefore, 553.

Further statistics in FD show that almost 45% of the patients (n=248) have at least one DOS score equal to or higher than three (which gives a positive indication for delirium), whilst never receiving the clinical diagnosis of delirium at any point during their stay. These findings strongly endorse our earlier findings of low test specificity of DOS. Using CAM, just four patients obtained a positive result but did not develop delirium during their stay.

The subsets with positive delirium patients and false-positive DOS patients are significantly smaller. The mean number of events per case and mean length of stay differ widely from the entire dataset. Data characteristics of the full dataset and the two subsets are shown in Table 1.

Table 1. Data characteristics of the patients in the full dataset and both subsets

	Full dataset (n=553)	Subset delirium positive (n=37)	Subset continuous DOS false-positive (n=78)
Delirium positive	37	37	0
Unique interventions	430	161	189
Total events	159,621	22,171	35,731
Mean events per case	289	599	458
Mean length of stay (days)	22.3	27.2	18.9
Positive DOS results	793	140	240
Negative DOS results	1,341	45	0
Positive CAM results	36	32	2
Negative CAM results	1,980	96	230

3.2. Process Models

A part of the process model built with the complete dataset is shown in Figure 1. This is a small extract of the entire model, which is too comprehensive to display in its entirety. In this process model, 20% of all activities are shown, as including more activities only made the model more uninterpretable. The percentage of paths shown in Figure 1 is 0%. Frankly, this does not mean that no paths are shown, but it is the absolute minimum amount possible in FD. Setting this percentage higher resulted in an even harder to read process model. Because of the low inclusion of activities and paths in this model, it only shows the most common cases and care paths and discards anomalies. Nevertheless, it is too wide-ranging to be readable. It does emphasise the complexity of care for delirium patients. It gives a good overview of the most commonly followed care paths and occurring interventions, but cannot clarify the emergence of delirium or the deviant diagnostic test results.

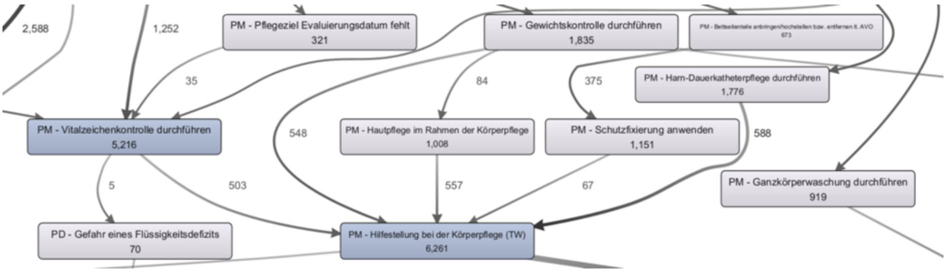

Figure 1. Extract from the process model built with the complete data set. The activities and paths thresholds are set to 20% and 0%, respectively.

The second visualisation only encompasses the positive delirium patients. We set the thresholds to show 30% of activities and the minimum for the paths. Although indistinguishable, this model gave a good overview of delirium patients' most common nursing interventions and DOS and CAM scores. The colours of the activity boxes and the thickness of the paths combined with the real-time animation of the process model highlight the most frequent events and paths patients encounter. Furthermore, it shows that the nursing diagnoses ‘acute confusion’ and ‘chronic confusion’ are often set relatively short before or after the clinical diagnoses of delirium and the DOS and CAM scores (n=17 and n=10). Because these conditions have many similarities with delirium, this could cause false positives in the detection questionnaires. Unfortunately, with these settings, we cannot derive information on the process before the DOS and CAM scores are set and why these are deviant because not all paths and activities are shown. Increasing this amount did not help to improve the understandability as the model became impossible to interpret due to the resulting broadened complexity.

The third visualisation is expansive and contains a wide variety of activities. We now set the activities threshold to 40% whilst maintaining the same level for the path threshold, resulting in a model with visual similarities to the previous. It highlights the most common paths and activities but cannot present an average or abnormal care process. The animation shows the flow of a patient through the care process correctly, which seems livelier and more diverse than the flow in the previous subset but has no additional value in comprehending the care process. The biggest problem with the animation in both models is that it does not show the patient’s state of DOS and CAM scores and diagnoses during the process. The scores and diagnosis are recorded as events in the dataset whilst applicable for one week in the real-world process.

3.3. Combat complexity

We tried filtering out nursing interventions unrelated to delirium care, such as ‘taking patient weight’ or ‘oral hygiene measures’. Even when omitting this data, it was hard to distil events to explain the low detection accuracy. Models with this filtered data show the same high frequent activities we saw in the subsets of patients but do not add anything valuable. Furthermore, omitting activities from the process model is only possible to a certain extent because of the variable character of delirium care.

4. Discussion

We automatically built process models using routinely collected nursing data from two gerontopsychiatric wards. The data encompassed one year of care and included over 700 patients. We made two subsets of the data to build the models, focusing on delirium positive and DOS false-positive patients. We sought to use these models to explain the low test accuracy of the DOS and CAM delirium detection instruments. Because of the wide variety of the data and subsequent breadth of the process models, they were of little use to better understand the process and answer our questions on the test accuracy.

4.1. Data Limitations

We found four limitations in our dataset whilst analysing the process models. These limitations made the models hard to interpret and detracted their explanatory value of the care process. The limitations apply to our data but can be applied to all types of routine healthcare data.

- Recorded events have a high level of variety.
- The real intervention time can differ from the registered intervention time.
- Each patient has a unique care path.
- Data from different healthcare professions is stored in different documentation systems, and one source may not contain enough data to see the whole story.

The first limitation is evident in our dataset, with 430 unique activities. Although this is in our study caused by the variability of delirium and gerontopsychiatric care, this applies to real-world healthcare data in general. It is heterogenic because conditions express themselves differently in each patient, and interventions are tethered to be as effective as possible in each case. This is, for example, emphasised by the International Classification of Health Interventions, which contains approximately 7,000 unique healthcare interventions [17]. To make meaningful use of process mining on healthcare data, it could be applied to datasets focused on specific care paths or after thorough data preparation to reduce the number of unique events and thus the complexity.

The time difference between the intervention and registration time expresses itself in our data because nurses work through a checklist when caring for a patient. Most of the time, they only really tick the boxes in the system after they visit the patient. This is, of course, coherent to focus on the patient whilst providing care but affects the representation of the process model. Unfortunately, this is hard to combat in routine clinical data. On top of that, our data is anonymised by using relative timestamps. This makes it impossible to track back if interventions happened, for example, during the night, which could be significant in identifying patterns.

The third limitation is well-known in healthcare and has been an essential topic for policymakers for years. Each patient's unique care path causes trouble when applying process mining and makes it harder to plan and evaluate care and predict outcomes. One way of dealing with this problem is to use clinical pathways; tools to guide evidence-based healthcare and translate guidelines into processes that standardise care for a specific intervention or condition. However, a recent literature review shows that although making care more (cost-)effective, the implementation of clinical pathways is a complex process and evidence on successful use is still sparse [18].

Finally, the last limitation is another problem that has been evident in healthcare for decades. Patient data is often scattered across many 'silos' because each department, and even each healthcare profession, uses its own software products. This dissemination makes it hard to retrieve all data from one patient and get a complete overview of the care path. In this case, we only used data from the nursing documentation. Especially in process mining, where a deviation from the standard nursing process can be caused by, for example, laboratory findings or medication information, it is essential to get all patient data.

4.2. Relation to other work

Although less specific, related limitations are also identified in other process mining studies. A literature review on the use of process mining in primary care identified several challenges, which included "data quality and how to assess or correct for consistency and completeness of routinely collected data" and "how to give insight into the discovered processes either by improved visualisations or comprehendible models" [19]. This emphasises the difficulty of working with routinely collected data and interpreting its models. Another study identified problems in their dataset regarding the completeness of the data [20]. Furthermore, all dates in their dataset had been shifted to a random date in the future to guarantee anonymity, which they identified as an essential data quality issue. The most recent literature study on process mining in healthcare resulted in many characteristics and challenges [21]. All limitations we identified are also present in their research. However, the limitations in our study are based on a real-world case study rather than a literature review. Furthermore, we focused specifically on routinely collected nursing data, which is unique in process mining to date.

4.3. Implications for Future Research

These limitations and associated recommendations can be used in further research to prepare data for process mining. It could prevent the overestimation of the applicability of process models made with routine medical data in the future. Furthermore, we found notable statistics on the outcomes of especially the DOS instrument. Although we could not find an explanation for these outcomes, it could be a significant reference point for subsequent studies on the use of the DOS in gerontopsychiatric patients.

Even with these four limitations, we do recognise the potential of process mining in healthcare. The technique is up-and-coming and can make meaningful re-use of routinely collected data. However, before routine medical data can be used to build a valuable process model, some preconditions should be met. This can help make the data more suitable for inclusion in a process model. Nevertheless, not all real-world data has the required properties to build a process model, either because of the overhead process or the quality of the data itself.

Although we did not meet our principal research goal, we got a clear overview of process mining in healthcare. We identified four limitations in our dataset, which should be kept in mind when mining all sorts of processes from routine medical data. We will continue working with this dataset and use our findings from this study to improve our process models in the future.

References

[1] Data Age 2025. The Digitization of the World From Edge to Core (22), IDC, USA, 2018.
[2] C. Sinsky et al., Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties, *Ann Intern Med* **165**(11) (2016), 753-760.
[3] M. Chernew and H. Mintz, Administrative Expenses in the US Health Care System: Why So High?, *JAMA* **326**(17) (2021), 1679–1680.
[4] M. Ghasemi and D. Amyot, Process mining in healthcare: a systematised literature review, *Int. J. Electronic Healthcare* **9**(1) (2016), 60–88.
[5] W. Aalst van der, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer-Verlag, Berlin, 2011.
[6] R. Mans et al., Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital, *Communications in Computer and Information Science* **25** (2008), 425-438.
[7] E. Reynish et al., Epidemiology and outcomes of people with dementia, delirium, and unspecified cognitive impairment in the general hospital: prospective cohort study of 10,014 admissions, *BMC Med* **15** (2017), 140.
[8] M. Schuurmans et al., The Delirium Observation Screening Scale: a screening instrument for delirium, *Research and theory for nursing practice* **17**(1) (2003), 31-50.
[9] S. Inouye et al., Clarifying Confusion: The Confusion Assessment Method, *Annals of Internal Medicine*. **113** (1990), 941-948.
[10] K. Gavinski et al., Validation of the delirium observation screening scale in a hospitalized older population, *Journal of hospital medicine* **11**(7) (2016), 494–497.
[11] Q. Shi et al., Confusion assessment method: a systematic review and meta-analysis of diagnostic accuracy, *Neuropsychiatr Dis Treat* **9** (2013), 1359–1370.
[12] M. Corrada et al., Dementia incidence continues to increase with age in the oldest old: the 90+ study. *Ann Neurol* **67**(1) (2010), 114-121.
[13] E. Stordal et al., The association between age and depression in the general population: a multivariate examination, *Acta Psychiatr Scand* **107**(3) (2003), 132-141.
[14] M. Netzer et al., Evaluating Performance and Interpretability of Machine Learning Methods for Predicting Delirium in Gerontopsychiatric Patients, Studies in HTI, **271** (2020), 121-128.
[15] W. Hackl et al., Visual Analytics in Delirium Management, *Studies in Health Technology and Informatics*, **279** (2021), 147–148.
[16] Fluxicon, Process Mining and Process Analysis, https://fluxicon.com/, last access: 7.2.2022.
[17] ICHI Beta-3 Reference Guide, ICHI Task Force, Swiss, 2020.
[18] T. Rotter, Clinical pathways as a quality strategy, in: Improving healthcare quality in Europe: Characteristics, effectiveness and implementation of different strategies. *European Observatory on Health Systems and Policies*, Copenhagen, 2019. pp. 309-330.
[19] R. Williams et al., Process Mining in Primary Care: A Literature Review, *Studies in health technology and informatics*, **247** (2018), 376-380.
[20] A. Kurniati et al., The assessment of data quality issues for process mining in healthcare using Medical Information Mart for Intensive Care III, a freely available e-health record database, *Health Informatics Journal*, **25**(4) (2019), 1878-1893.
[21] J. Munoz-Gama et al., Process Mining for Healthcare: Characteristics and Challenges, *Journal of Biomedical Informatics*, 2022.