Modeling Patient Trajectories in a Quebec Health Institution: Developing a Process Mining-Based Decision Support Tool for Healthcare Management

Alexandra Langford-Avelar[1,2,3,*]

[1]School of Public Health, Université de Montréal, Quebec, Canada
[2]Centre de recherche en santé publique (CReSP), Université de Montréal and Centre intégré universitaire de santé et de services sociaux du Centre-Sud-de-l'Île-de-Montréal, Quebec, Canada
[3]Unité de Soutien SSA Québec, Quebec, Canada

Abstract
Healthcare institutions face growing challenges in ensuring care continuity across increasingly complex service networks. This doctoral project focuses on designing a process mining-based tool to visualize and analyze patient trajectories across the continuum of care, supporting healthcare managers in operational decision-making within a Quebec institution. The tool will support healthcare managers by generating interpretable process maps from large-scale administrative health data, enabling improved decision-making and continuity of care. The project unfolds in three phases: (1) centralizing and standardizing data; (2) applying and refining process mining techniques to generate multi-level trajectory models from individual patient logs; and (3) integrating the algorithm into an interactive tool co-developed with institutional partners. This research contributes to a novel, system-wide approach to understanding care processes and aligns with the development of learning health systems.

Keywords
Process mining, Patient trajectories, Health information system, Administrative health data, Healthcare continuity, Decision support tool

1. Background

Healthcare institutions face challenges in ensuring care coordination and continuity, particularly in supporting well-integrated patient's journeys. Although vast amounts of health data are generated daily, much of it remains underused for gaining a comprehensive understanding of healthcare trajectories. This underutilization hampers informed decision-making and may contribute to inefficiencies across healthcare systems [1, 2, 3].

This research addresses a gap in current healthcare management practices: the lack of a comprehensive, data-driven, and adaptive mechanism to accurately model and optimize healthcare trajectories across fragmented information systems. The motivation for this project stems directly from the persistent challenges in achieving true care coordination and continuity, which are fundamental to patient-centered care [1, 2, 3]. The traditional reliance on static, predefined clinical pathways fail to capture the dynamic, iterative, and often non-linear nature of patient's journeys [4]. This deficiency not only limits our understanding of patient's needs but also leads to suboptimal care, increased costs, compromised health outcomes, and diminished provider well-being.

The proposed solution adopts a process-based approach to model patient's trajectories using administrative health data. Unlike prescriptive models, this approach reflects how care unfolds in time and space, acknowledging the inherent complexity and variability of healthcare trajectories [5, 6, 7, 8]. It enables the verification of real-world patient pathways against established clinical guidelines, highlighting deviations and inefficiencies [9, 10]. Viewing each patient healthcare trajectory as a complex, evolving

ICPM Doctoral Consortium and Demo Track 2025, October 20-24, 2025, Montevideo, Uruguay
*Corresponding author.
alexandra.langford.avelar@umontreal.ca (A. Langford-Avelar)
0000-0002-4651-618X (A. Langford-Avelar)

system fosters a more comprehensive, flexible, and actionable understanding of care processes. This shift from "what should happen" to "what actually happens" is fundamental to proactive management and resource allocation [11].

The importance of this project is particularly pronounced within the Quebec healthcare context. Despite the evident advantages of interoperable health information systems (HIS), Quebec's health and social services network is characterized by significant fragmentation, with over 9,000 distinct HIS identified in 2019 [12]. These systems, often designed for specific local needs, create pervasive data silos, both between and within a healthcare institution [12]. As patients interact with various facilities (e.g., ER, laboratories, long-term care), their administrative health data are dispersed across these specialized, disparate systems. The absence of a unique patient identifier, coupled with multiple data formats and a lack of standardized data entry practices, exacerbates this fragmentation, making a holistic view of patient's journeys nearly impossible [13].

This project aims to develop a decision support tool that uses process mining to generate interpretable models of patient trajectories from fragmented administrative health data within a Quebec healthcare institution. Instead of focusing on specific diagnoses or isolated departmental workflows, the tool will offer a comprehensive view of service utilization across the entire continuum of care. By consolidating and analyzing dispersed data sources, the project seeks to overcome interoperability barriers and produce actionable insights into actual care processes. By addressing common challenges in data fragmentation and care coordination, the tool offers a transferable framework that can be deployed across diverse healthcare systems. This integrated approach will support healthcare managers in making informed, data-driven decisions and ultimately contribute to enhancing the quality and efficiency of care delivery.

2. Objective

This project focuses on designing a process mining-based framework to model and visualize healthcare trajectories from administrative data, supporting operational decision-making within a major Quebec healthcare institution. Co-developed with the CIUSSS de l'Ouest-de-l'Île-de-Montréal (COMTL) this project leverages six years (2020-2026) of anonymized administrative health data from multiple care settings (e.g., hospitals, long-term care, outpatient clinics) sourced from 45 distinct information systems.

To achieve this main objective, three sub-objectives are required:

1. Centralize COMTL's administrative health data into a standardized database. *(completed)*
2. Apply and refine process mining techniques to generate aggregated representations of patient trajectories from individual logs, ensuring the models are both methodologically robust and operationally relevant for COMTL decision-makers.
3. Integrate the algorithm into a functional tool and validate its decision-support potential through case studies aligned with COMTL's management priorities.

The project includes a methodological development phase aimed at designing and testing algorithms to meet the three sub-objectives using historical data (2020 to 2024). This phase will validate the approach while developing and refining the algorithms. Once validated, these algorithms will be deployed to continuously integrate new data (2025 to 2026). The resulting tool will enable the visualization of COMTL healthcare trajectories across different time frames, levels of granularity, and population strata.

3. Methods

3.1. Objective 1: Creation of standardized databases *(completed)*

The first objective was the creation of two centralised and standardized databases: one dedicated to patients and another to the services they received. This was achieved using Extract, Transform, Load (ETL) processes implemented in SQL. The completion of this objective provides the essential structured data necessary for the subsequent analysis and modeling of patient care trajectories.

Patient Database

Fragmentation across healthcare information systems creates multiple administrative records for the same patient, each identified differently across facilities, services, and databases. This fragmentation limits the ability to track a patient's complete journey, posing a major barrier to ensuring continuity of care [14]. To address this, we developed a deterministic matching algorithm that consolidates patient records into a unique identifier [15]. The algorithm leverages available socio-demographic data and links all relevant record numbers, using a robust matching key based on the structure of the health insurance number (Figure 1A). When discrepancies appeared across records, we applied tailored rules depending on the variable type. For time-dependent variables like addresses or phone numbers, the algorithm consistently retained the most recent data, thereby reflecting the evolution of information over time. For fixed variables, such as name or date of birth, the agreement across different sources was evaluated using contingency tables. This process enabled the consolidation of approximately **10 million active records** (2020-2024) into **3 million unique patients**. From this consolidated dataset, we were able to assess the consistency of key attributes, such as sex, date of birth, and first and last names, across the various information systems that contributed to our data (Figure 1B).

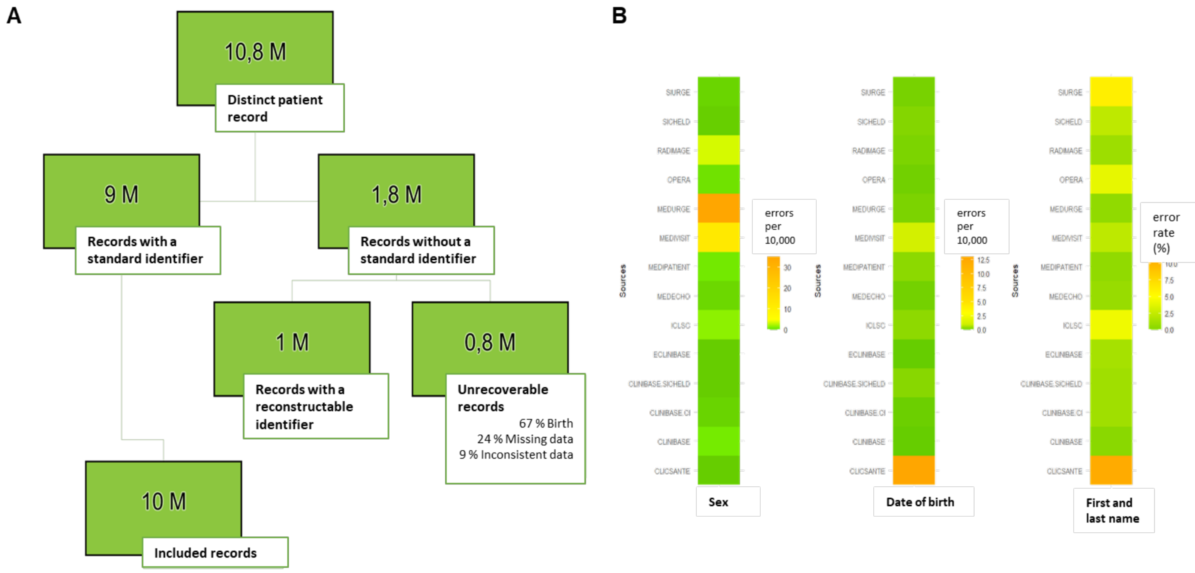

Figure 1: Creation of a single, standardized patient index
A) Patient record selection filter ; B) Modality concordance assessment.

Service Database

The second database, the service database, was constructed to capture all healthcare services provided across the various COMTL facilities. This was achieved through a comprehensive union of service data extracted from the diverse available sources. The pivotal step involved linking each service record to the unique patient key that had been generated in the "patient database". This linkage enabled the aggregation of all services associated to a single individual, thereby laying the groundwork for the generation of both individual and aggregated patient trajectories.

3.2. Objective 2: Aggregation strategies and process mining techniques

This objective will leverage the previously established service database, transforming raw service data into actionable insights through the application of process mining. The initial step involves creating an event log including unique patient identifiers, specific activities performed, and precise timestamps (start and end). To reduce model complexity and avoid the so-called "spaghetti effect" associated with large-scale data, the analysis will implement two levels of data aggregation: a macro level for **main activities** (e.g., hospitalization) and a detailed level for associated **secondary activities** (e.g., radiology during hospitalization). This dual-granularity approach will preserve data richness while ensuring interpretability across service, facility, or institutional levels.

The generated event log will then serve as input for the modeling algorithm, which will generate process maps as graphs, with nodes representing activities and arcs indicating their chronological sequence. To ensure methodological rigor and build upon existing research, this phase will begin with a comparative evaluation of established process mining algorithms. These techniques will be tested on the constructed event logs to assess their suitability for modeling complex and heterogeneous patient trajectories. Descriptive statistics, such as activity volumes, durations, and proportions, will be integrated into both nodes and arcs. A key focus is managing the complexity of process maps. To address this, we will apply strategies such as filtering low-frequency arcs using the Pareto principle (retaining 80% of significant trajectories, with filtered arcs analyzed independently for deviations) and implementing hierarchical aggregation to enable flexible visualization across different organizational levels [16].

The quality of the process maps will be evaluated against four interdependent criteria: Fitness (model's ability to reproduce trajectories), Precision (proposing only observed paths), Generalization (considering valid but unobserved trajectories), and Simplicity (clear interpretation). Striking the right balance between precision and generalization will require careful calibration, with validation from clinical experts to avoid producing inconsistent or misleading representations.

3.3. Objective 3: Tool integration and collaborative validation for decision support

The third objective focuses on integrating the developed algorithms into an operational, interactive dashboard designed to support decision-making by healthcare managers at COMTL. This tool will be developed using Power BI, a software already widely adopted within Quebec's healthcare network, ensuring compatibility with existing workflows and user familiarity. The dashboard will enable dynamic visualization of process maps and key indicators through customizable filters (e.g., temporal window, patient strata, level of granularity), allowing users to explore trajectories across various dimensions. A central aspect of this objective lies in the close collaboration with COMTL managers, clinical leads, and field experts. Their involvement will be instrumental in validating the interpretability and operational relevance of the generated process maps. Through a series of co-development workshops and iterative feedback cycles, the tool's functionalities and outputs will be aligned with the institution's operational needs. This approach ensures that the trajectories generated by the algorithm meaningfully address current management challenges, such as care delays, continuity disruptions, and resource bottlenecks.

Beyond its descriptive capabilities, the tool will include a proactive alert system designed to flag atypical situations requiring managerial attention. This includes both trajectory-level anomalies, such as care sequences or durations that deviate significantly from established patterns, and volume-level alerts, such as sudden increases in patient flow that may exceed the institution's available resources (e.g., exceeding bed capacity or surpassing scheduled staff availability). These alerts will be configured in collaboration with domain experts, based on thresholds reflective of clinical or organizational priorities. Such cases may signal quality or coordination issues and will be subject to in-depth investigation to identify potential inefficiencies or risks.

4. Contributions and Perspectives

This research builds upon and extends existing work in process mining, particularly within healthcare. While process mining has been applied to clinical guideline compliance and departmental analysis, few studies address its use in large-scale governance or service integration in healthcare systems. The research draws from foundational work and adds methodological innovation in managing scale and complexity, resolving interoperability challenges across fragmented information systems, and ensuring real-time actionable use of process models. By embedding process models into a local governance tool used by managers and clinicians, this project bridges the gap between data science research and health system transformation. It actively contributes to the development of a Learning Health System

(LHS), where health data are continuously harnessed to monitor, adapt, and improve service delivery across the care continuum. In addition to supporting operational decision-making, the tool provides a methodological framework that can be reused across diverse healthcare systems.

5. Timeline

The timing of this submission is particularly well-aligned with the progression of my project. I have recently completed and successfully defended my research protocol, receiving positive feedback from field experts. In parallel, I have achieved the first objective of the project: the development of centralized databases. As I now enter the algorithm development phase for process mining, this doctoral consortium offers an ideal opportunity to present tangible progress while remaining receptive to expert feedback to further refine and strengthen the methodological approach.

6. Declaration on Generative AI

During the preparation of this work the author used DeepL in order to improve language and readability. After using this tool, the author reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] J. Haggerty, R. Reid, G. Freeman, B. Starfield, C. Adair, R. McKendry, Continuity of care: a multidisciplinary review, BMJ 327 (2003 Nov 22). URL: https://www.bmj.com/content/327/7425/1219.long. doi:10.1136/bmj.327.7425.1219.

[2] A. Katz, Focus on continuity of care, Can Fam Physician 68 (2022) 483. doi:10.46747/cfp.6807483_2.

[3] R. Khatri, A. Endalamaw, D. Erku, E. Wolka, F. Nigatu, A. Zewdie, Y. Assefa, Continuity and care coordination of primary health care: a scoping review, BMC Health Serv Res 23 (2023) 750. doi:10.1186/s12913-023-09718-8.

[4] G. Mould, J. Bowers, M. Ghattas, The evolution of the pathway and its role in improving patient care, Quality safety in health care 19 (2010 Oct). URL: https://pubmed.ncbi.nlm.nih.gov/20430932/. doi:10.1136/qshc.2009.032961.

[5] A. Kshenin, S. Kovalchuk, Data-driven modeling of complex business process in heterogeneous environment of healthcare organization with health information systems, pHealth 2021 (2021). URL: https://ebooks.iospress.nl/doi/10.3233/SHTI210583. doi:10.3233/SHTI210583.

[6] R. Lenz, M. Peleg, M. Reichert, Healthcare process support: Achievements, challenges, current research, International Journal of Knowledge-Based Organizations 2 (2012).

[7] P. Harper, A framework for operational modelling of hospital resources, Health care management science 5 (2002 Aug). URL: https://pubmed.ncbi.nlm.nih.gov/12363044/. doi:10.1023/a:1019767900627.

[8] D. Schuster, E. Benevento, D. Aloini, W. M. P. van der Aalst, Analyzing healthcare processes with incremental process discovery: Practical insights from a real-world application, Journal of Healthcare Informatics Research 2024 8:3 8 (2024 Jun 24). URL: https://link.springer.com/article/10.1007/s41666-024-00165-6. doi:10.1007/s41666-024-00165-6.

[9] R. Mans, W. M. P. van der Aalst, R. Vanwersch, Process mining in healthcare, SpringerBriefs in Business Process Management (2015). URL: https://link.springer.com/book/10.1007/978-3-319-16071-9. doi:10.1007/978-3-319-16071-9.

[10] J. Munoz-Gama, N. Martin, C. Fernandez-Llatas, O. A. Johnson, M. Sepúlveda, E. Helm, V. Galvez-Yanjari, E. Rojas, A. Martinez-Millana, D. Aloini, I. A. Amantea, R. Andrews, M. Arias, I. Beerepoot, E. Benevento, A. Burattin, D. Capurro, J. Carmona, M. Comuzzi, B. Dalmas, R. de la Fuente, C. Di Francescomarino, C. Di Ciccio, R. Gatta, C. Ghidini, F. Gonzalez-Lopez, G. Ibanez-Sanchez,

H. B. Klasky, A. Prima Kurniati, X. Lu, F. Mannhardt, R. Mans, M. Marcos, R. Medeiros de Carvalho, M. Pegoraro, S. K. Poon, L. Pufahl, H. A. Reijers, S. Remy, S. Rinderle-Ma, L. Sacchi, F. Seoane, M. Song, A. Stefanini, E. Sulis, A. H. M. Ter Hofstede, P. J. Toussaint, V. Traver, Z. Valero-Ramon, I. V. Weerd, W. M. P. van der Aalst, R. Vanwersch, M. Weske, M. T. Wynn, F. Zerbato, Process mining for healthcare: Characteristics and challenges, Journal of Biomedical Informatics 127 (2022) 103994. doi:10.1016/j.jbi.2022.103994.

[11] F. Ruiz, F. Garcia, L. Calahorra, C. Llorente, L. Gonçalves, C. Daniel, B. Blobel, Business process modeling in healthcare, Studies in Health Technology and Informatics 179 (2012) 75–87.

[12] G. du Québec., Plan de transformation numérique 2023-2027 : secteur de la santé et des services sociaux version finale, 2023. URL: https://intranetreseau.rtss.qc.ca/index.php?accueil.

[13] A. Torab-Miandoab, T. Samad-Soltani, A. Jodati, P. Rezaei-Hachesu, A. Torab-Miandoab, T. Samad-Soltani, A. Jodati, P. Rezaei-Hachesu, Interoperability of heterogeneous health information systems: a systematic literature review, BMC Medical Informatics and Decision Making 2023 23:1 23 (2023). URL: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-023-02115-5. doi:10.1186/s12911-023-02115-5.

[14] M. T. Neame, J. Chacko, A. E. Surace, I. P. Sinha, D. B. Hawcutt, A systematic review of the effects of implementing clinical pathways supported by health information technologies, Journal of the American Medical Informatics Association 26 (2019) 356–363. doi:10.1093/jamia/ocy176.

[15] liklich. R.E., N. Dreyer, M. Leavy, Registries for Evaluating Patient Outcomes: A User's Guide, Managing Patient Identity Across Data Sources., 2014. URL: https://www.ncbi.nlm.nih.gov/books/NBK208618/.

[16] D. Bosson-Rieutort, A. Langford-Avelar, J. Duc, B. Dalmas, Healthcare trajectories of aging individuals during their last year of life: application of process mining methods to administrative health databases, BMC Medical Informatics and Decision Making 25 (2025) 58. doi:10.1186/s12911-025-02898-9.