BMC Medical Research Methodology

RESEARCH

Open Access

Toward value-based care using cost mining: cost aggregation and visualization across the entire colorectal cancer patient pathway

Maura Leusder[1†], Sven Relijveld[2,3†], Derya Demirtas[2,4], Jon Emery[3], Michelle Tew[5], Peter Gibbs[3,6], Jeremy Millar[7,8], Victoria White[9], Michael Jefford[10], Fanny Franchini[3,5†] and Maarten IJzerman[1,2,3,5,10*†]

Abstract

Background The aim of this study is to develop a method we call "cost mining" to unravel cost variation and identify cost drivers by modelling integrated patient pathways from primary care to the palliative care setting. This approach fills an urgent need to quantify financial strains on healthcare systems, particularly for colorectal cancer, which is the most expensive cancer in Australia, and the second most expensive cancer globally.

Methods We developed and published a customized algorithm that dynamically estimates and visualizes the mean, minimum, and total costs of care at the patient level, by aggregating activity-based healthcare system costs (e.g. DRGs) across integrated pathways. This extends traditional process mining approaches by making the resulting process maps actionable and informative and by displaying cost estimates. We demonstrate the method by constructing a unique dataset of colorectal cancer pathways in Victoria, Australia, using records of primary care, diagnosis, hospital admission and chemotherapy, medication, health system costs, and life events to create integrated colorectal cancer patient pathways from 2012 to 2020.

Results Cost mining with the algorithm enabled exploration of costly integrated pathways, i.e. drilling down in high-cost pathways to discover cost drivers, for 4246 cases covering approx. 4 million care activities. Per-patient CRC pathway costs ranged from $10,379 AUD to $41,643 AUD, and varied significantly per cancer stage such that e.g. chemotherapy costs in one cancer stage are different to the same chemotherapy regimen in a different stage. Admitted episodes were most costly, representing 93.34% or $56.6 M AUD of the total healthcare system costs covered in the sample.

Conclusions Cost mining can supplement other health economic methods by providing contextual, sequence and timing-related information depicting how patients flow through complex care pathways. This approach can also facilitate health economic studies informing decision-makers on where to target care improvement or to evaluate the consequences of new treatments or care delivery interventions. Through this study we provide an approach

[†]Maura Leusder and Sven Relijveld contributed equally to this work.

[†]Fanny Franchini and Maarten IJzerman contributed equally to this work.

*Correspondence:
Maarten IJzerman
ijzerman@eshpm.eur.nl
Full list of author information is available at the end of the article

for hospitals and policymakers to leverage their health data infrastructure and to enable real time patient level cost mining.

Keywords Costs of care, Colorectal cancer, Patient pathways, Process mining, Value-based healthcare

Introduction

Recent years have witnessed significant advancements in complex care, particularly in oncology, with rapid introduction of innovative technologies and therapies. This has led to better patient outcomes but has also resulted in higher patient-specific costs due to increased complexity and specialization of care delivery [1, 2]. Recent estimates suggest that the total global economic burden of cancers will reach $25.2 trillion during the period of 2020 to 2050 [3]. This rapidly growing cost of care is unsustainable and considered one of the major challenges for health systems worldwide [2]. Value-based healthcare (VBHC) is a lens through which this issue is increasingly discussed; broadly speaking, VBHC suggests that healthcare must be organized and incentivized in a way that prioritizes outcomes and minimizes resource utilization and costs, per patient, across the integrated treatment pathway from screening or initial consultation to outcome [4]. While patient preferences and outcomes are increasingly studied, estimating costs at the patient level remains challenging [4], especially in complex care settings with extended patient journeys or repetitive treatment cycles with regular diagnostic work-ups, such as colorectal cancers (CRC). As new treatment variations and alternatives are introduced, and protocols become more tailored to individual patients, these pathways increasingly resemble interdependent webs which complicates decision-making [5–8].

Model-based health economic studies often use population-level aggregate costs and rely on ad-hoc exploration of variability or cost drivers within these aggregates, usually based on patient characteristics like age [9–12]. While suitable for evaluating interventions, this approach is less accurate for hospital-level capacity planning and process improvement [13–18]. Additionally, healthcare professionals report a lack of tools to easily identify and target specific cost drivers relevant to their local context [10, 18–20]. Determining cost drivers across patient pathways is a significant research challenge [3, 21–23], as decisions made in one treatment impact subsequent treatments' costs and outcomes, prompting calls for better tools to systematically explore variation across integrated pathways [5, 8, 18, 24–27]. Granular cost data spanning the full patient cycle, from primary care to end-of-life care, are difficult to generate [4, 28, 29], and determining variation in healthcare delivery characteristics remains a core challenge.

To address these challenges, this study presents process mining with cost estimation, which we call "cost mining," as an approach to uncover high-cost pathways and specific cost drivers using real-world patient-level data. Process mining (PM) can complement existing health economic approaches [13, 30], by enabling patient-level cost estimates in models and generating visuals that capture patient-level variation and treatment interdependencies. PM uses low-level event data from electronic health records (EHR), such as individual consultations, procedures, and medication prescriptions, with timestamps to derive process models and discover real-world patient pathways [31]. It presents granular data in steps or phases, providing descriptive insights into patient movement through systems and resource consumption [31, 32]. As of early 2022, approximately 263 healthcare PM studies have been published [30], exploring care trajectories in acute ischemic stroke, sepsis [33], chronic diseases [34, 35], cancer [36–38], primary care [32], and COVID-19 cases [28]. This work has concluded that PM is powerful, but should include cost or resource data to make it actionable, which is what we aim to contribute in this study.

Costs have received limited attention in prior PM and VBHC studies. PM has been used to assess resource requirements and queuing improvements in emergency departments [14, 15, 18, 39], but its use in cancer care is limited due to the complexity of tracing integrated care episodes and the chronic nature of cancer [21, 22]. To support case-mix group evaluations and hospital capacity planning, additional data and analyses are needed with PM [14–16]. Cost mining can identify patient subgroups incurring additional costs due to factors like cancer stage, treatment timing, or protocol changes. It complements existing health economic methods by providing contextual information on patient pathways and the timing of treatment decisions (e.g., early-stage vs. late-stage chemotherapy). This information can serve as KPIs or benchmarks for healthcare practitioners, policymakers, and researchers, extending PM's usefulness in health services [30]. Given that only nine of 236 recently reviewed studies employed cost estimation [18, 24, 25, 27, 30], the algorithm we have developed particularly enhances PM's utility for studying the cost drivers in CRC and other complex diseases in scope for VBHC initiatives.

To develop and illustrate cost mining, we created a unique linked dataset to cover the integrated colorectal cancer (CRC) pathway in Victoria, Australia, which serves as an illustrative case study throughout the paper. Colorectal cancers, which have long trajectories beginning in primary care, are the most costly cancers in Australia [22] and the second most costly cancer globally [3], making CRC a highly relevant research context for the study of healthcare costs.

Methods

In this section we describe the data requirements for cost mining integrated pathways. For a detailed description of PM techniques, we refer the reader to Munoz-Gama et al. (2023) [31] and van der Aalst (2016) [40]. In this study, we combined data from six Australian databases, detailed in appendix A and summarized in Fig. 1. The study received ethical approval by the Royal Melbourne Hospital Ethics Board through the BioGrid application (202,003/8) prior to starting.

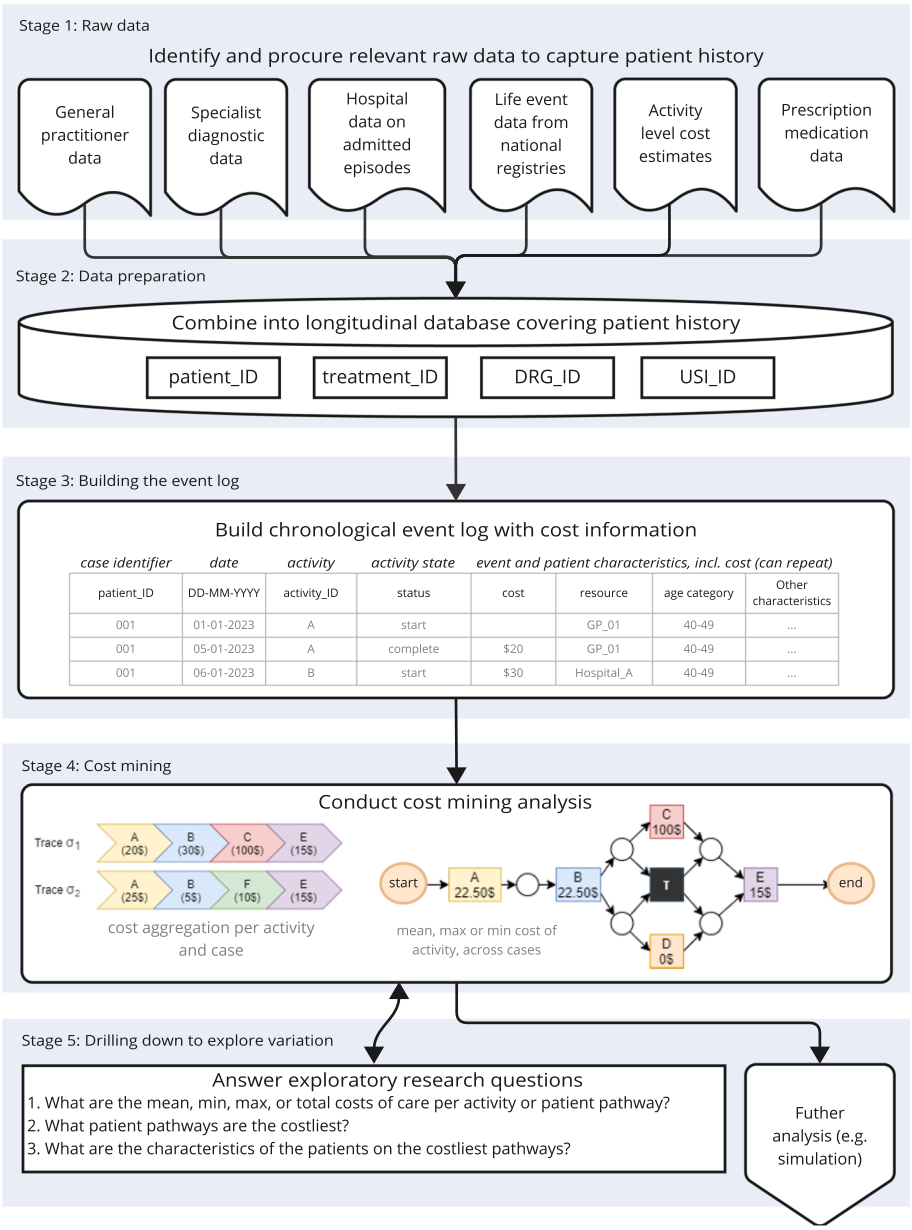


Fig. 1 Explanatory diagram summarizing the flow of raw data into research results in the proposed method using PM with cost aggregation

PM structures event-level data chronologically into so called process models, which depict a linear, visualized flow of patients through a series of processes [32, 40]. Processes can have several states and attributes (e.g. a blood test can be complete or incomplete, etc.). PM describes as-is states of pathways using retrospective data; it summarizes and visualizes real world pathways, and does not make any predictions, assumptions, or imputations [29, 32, 34, 41].

Stage 1: raw data

The method requires activity and cost information of a patient spanning the entire treatment history (screening, diagnosis, treatment, follow-up), and these activity data need to include dates or timestamps. Patients don't need to complete their treatment to be included in the analysis, as costs are estimated at the activity level, including patients still undergoing treatments is a key strength of this method. However, for group comparisons or total cost estimations, it is crucial to have treatment start dates to filter out incomplete cases and avoid downward bias in total pathway cost estimates [8]. Costs can be estimated using activity-based microcosting approaches [5, 8], or through reimbursement data such as DRGs [4, 12, 22]. The Australian reimbursements are granular, meaning that this method will produce cost statistics that capture inter-dependencies across integrated pathways. For example, the chemotherapy stage consists of several activity-based reimbursements, which means that the cost statistics will reflect differences between patients, as e.g. a patient requiring chemotherapy at a later stage of CRC may require more consultations, treatments, or regimens than a patient undergoing chemotherapy at a different CRC stage. The data requirements are summarized in the first stage of Fig. 1.

Stage 2: data preparation

The data need to be linked into a longitudinal database covering the integrated patient pathways and associated costs per activity. This implies that each data source identified in stage 1 of Fig. 1 needs to contain unique identifiers, e.g., anonymized patient identifiers. Further, it implies that data requirements are significant, because data linkage results in the exclusion of incomplete cases. In the CRC case shown in Fig. 2, this resulted in a set of 4246 patient records covering approximately 4 million activities (appendix A). Before conducting the analysis, it is important to assess if combining the data introduced bias through data loss, by comparing patient characteristics across data sources and the final set (appendix B).

Stage 3: building the event log

Next, data need to be formatted in an event or activity log, which is subject to the requirements summarized in Table 1.

An activity log contains one row per activity, with start and end times, and therefore only supports additional data at the unit of analysis of an activity as shown in Fig. 3. On the other hand, event logs offer more flexibility because they contain two or more rows per activity, as start and end points of activities are considered individual events [30, 40]. As such, it is possible to model data in which e.g. different resources are executing different elements of a single activity. A practical example of this would be a patient starting a medication-based treatment at a specialist care facility but completing it weeks later whilst being treated at a hospital for acute complications. For the purpose of cost mining, an event log is favorable to an activity log, because some healthcare activities can take weeks or months (e.g. medication treatment regimens), and others minutes (e.g. phone consultation) [30]. The largest challenge in PM in the healthcare sector is related to the inconsistent nature of the data required [30]. It can be challenging to link and combine data sources to cover integrated pathways in settings like CRC, due to the length or dispersion of treatments. Possible solutions for this include using heuristics to estimate process end times if these are unknown [8], or assuming that the start date of a specific activity signifies the end date of the prior one. In our CRC case, we did not make assumptions or imputations, because we constructed entire integrated care pathways from primary care up to outcomes like survivorship.

The event log should be built in software optimized for efficient coding, recoding, and reformatting of large data sets. We used R with the tidyverse library, which is freely available. The required event log format is shown in Fig. 3 exhibit A. Note that row 1 in the activity log contains the information from rows 1–2 in the event log. Further, note that the activity log in exhibit B loses some of the information contained in the event log (rows 3–4). The activity log cannot support data pertaining to an activity instance (start, end). Therefore, it summarizes the costs of activity B ($30) whereas the event log can show when and where these costs are incurred ($10 at start, $20 at completion).

Once the event (or activity) log is built as presented in the methods section (stage 1–3), the cost mining analysis can be conducted. Modern commercial PM software packages[1] support the display of common statistics, such

[1] https://www.fluxicon.com/disco (commercial).
https://www.celonis.com (commercial).
https://www.apromore.org (commercial).
https://www.promtools.org (free).

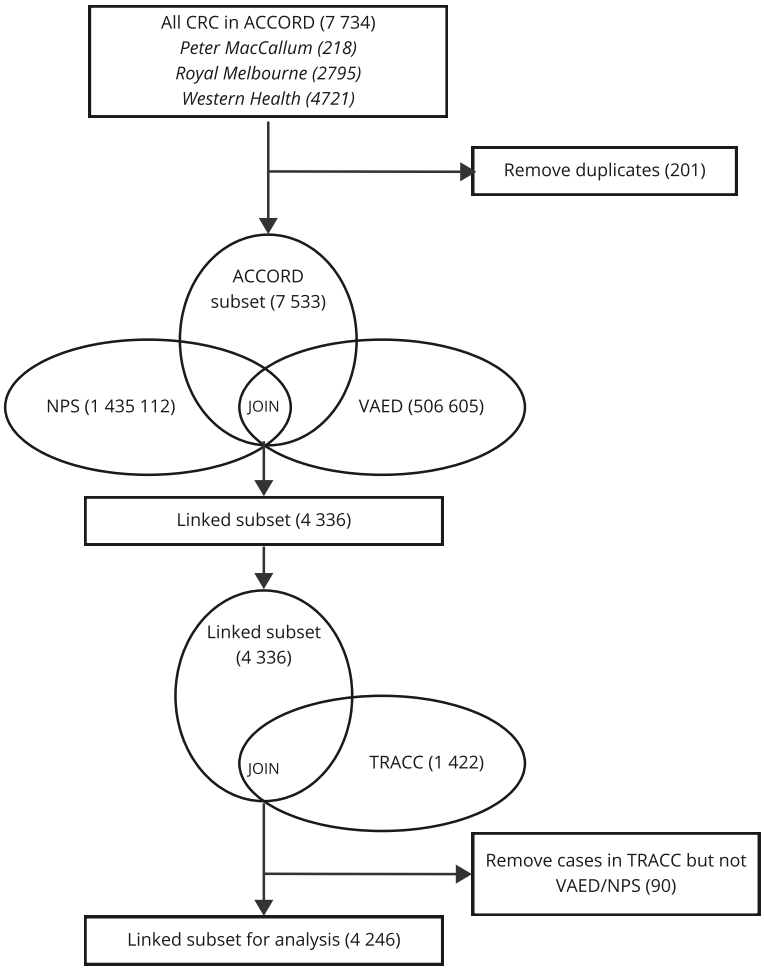


Fig. 2 Patient record selection for the illustrative case study of colorectal cancer, resulting in a dataset of 4,246 linked unique cases with cost data at the activity or event level, covering approx. 4 million activities. For details, please refer to appendix A. Note: ACCORD: Australian Comprehensive Cancer Outcomes and Research Database; MBS: Medicare Benefits Schedule; PBS: Pharmaceutical Benefit Scheme; TRACC: Treatment of Recurrent and Advanced Colorectal Cancer; VAED: Victorian Admitted Episodes Dataset

Table 1 Event log requirements, based on De Roock and Martin (2022) [30]

Element	Description
Timestamps	Dates, timestamps
Case identifier	A case identification code that is consistent and unique, e.g. one code per patient
Activity identifier	An activity identification code that is consistent and unique. This requires data cleaning and preparation to avoid cases where identical activities or events are coded inconsistently (e.g. "Chemo" vs. "Chemotherapy")
Event status	Activity status information, e.g. started, complete, in progress associated with the timestamps
Cost of event or activity	Cost estimates, stemming from e.g., diagnosis-related group codes or microcosting
Additional data	E.g. patient characteristics, case-mix group

as the median number of cases per activity, but do not support customized statistics such as cost information. For this reason, we wrote a customized cost mining algorithm in Python, which is used in the following analyses

Footnote 1 (continued)

https://pm4py.org (free for use in Python).
https://www.bupar.net (free library for use in R).

Exhibit A: Event log

	patient_ID	date	activity_ID	status	cost
1	001	01-01-2023	A	start	
2	001	05-01-2023	A	complete	$20
3	001	06-01-2023	B	start	$10
4	001	09-01-2023	B	complete	$20

Exhibit B: Activity log

	patient_ID	date start	date end	activity_ID	cost
1	001	01-01-2023	05-01-2023	A	$20
2	001	06-01-2023	09-01-2023	B	$30

Fig. 3 Minimum requirements of an event log or an activity log for PM with cost aggregation

(available https://github.com/chsr-uom/PM_token_decoration.)

Results

Stage 4: cost mining

The analysis starts with executing PM on the entire event log built in stage 3 using an inductive miner algorithm. It is particularly suitable to healthcare processes, because it produces inspectable process maps with a large degree simplification [32, 42–44]. Using the code we provide, the resulting process map displays cost statistics (mean, minimum, maximum, total) for each activity displayed in the form of a 'decoration' [45, 46], i.e. a label on the process map. For any given process model generated, the visual output provides the summary statistic of the costs per activity, based on the number of cases that have passed through the activity in that analysis. Similarly, it produces a summary statistic of the total costs of care per trace, i.e., per individual patient trajectory included. At this point, it can be useful to restrict the sample to cases that are completed to avoid under-estimating total pathway costs, by e.g. restricting the data to cases with an observed life event (e.g., survivorship, death, no treatment within 2 years). The cost mining code is described in pseudocode in appendix C. Figure 4 summarizes how the algorithm aggregates cost data; it draws on the traces derived from PM, which are sequences of events observed per case (patient) in the dataset. In simple terms, for each process map generated, the algorithm aligns all traces of the current model to calculate a statistic of the costs of each activity. In Fig. 4 exhibit B, both instances of 'activity A' are compared and translated into a mean (in this case, the average of $20 and $25 is $22.50). To do so, the algorithm accounts for all patients that have undergone activity A, across all traces (sequences of activities). Because, for example, only a single instance of activity C is observed in this hypothetical example, the label returns the value of $100 attached to activity C. In a final step, the code attaches the generated statistic value to the process map as a 'decoration' label [45, 46].

Stage 5: drilling down to explore variation

The generated process model will display pathways, which warrant further exploration in terms of e.g. case-mix groups, diagnoses, or indications, which we term 'drilling down' into the data to further understand rare, desirable, or undesirable pathways and cost drivers [30, 32, 40]. This allows us to quantify mean and range per patient group as well as to determine subgroups based on certain cost outcomes (e.g. most expensive).

We illustrate the method in Fig. 5 using the CRC case. We were able to identify crucial decision points (after which pathways were significantly different in complexity and costs), pinpoint costly processes, and make case-mix comparisons across groups (sex, age group, tumour location, tumour stage, CRC-type, patient's rurality, and indigenous status; see right side of Fig. 5). In CRC, we found that the average costs of care ranged from $10,379 AUD to $41,643 AUD per patient (Fig. 5 panel H) and differed significantly per stage of treatment.

Drilling down in our data revealed that colon cancer was associated with significantly greater costs across the entire care continuum than rectal cancer, and admissions and chemotherapy were by far the most expensive elements of treatment (Fig. 5, panels C, D). Admitted episodes ($n=1,965$ patients) cost a total of $56.6 M AUD (93.34% of total costs covered by the data, $ 60,63 M AUD). In comparison, the total cost of chemotherapy drug treatments ($n=218$ patients) was 6.62% of total costs. GP visits, diagnostic testing, and prescriptions made up less than 0.01% of the total costs. Our results reveal that treatment-related factors, namely cancer stage, significantly related to costs (Fig. 5, panel H).

When drilling down into the chemotherapy treatments, treatment with a specific regimen (Mfolfox 6; Fig. 5 panel D) was extremely costly, at an average cost of $35 K AUD per patient. However, these costs significantly varied across the different cancer stages, with stage C cancer patients incurring much higher costs associated with the Mfolfox 6 chemotherapy regimen than other patients, which warrants future qualitative and quantitative research. In this way, this exploratory

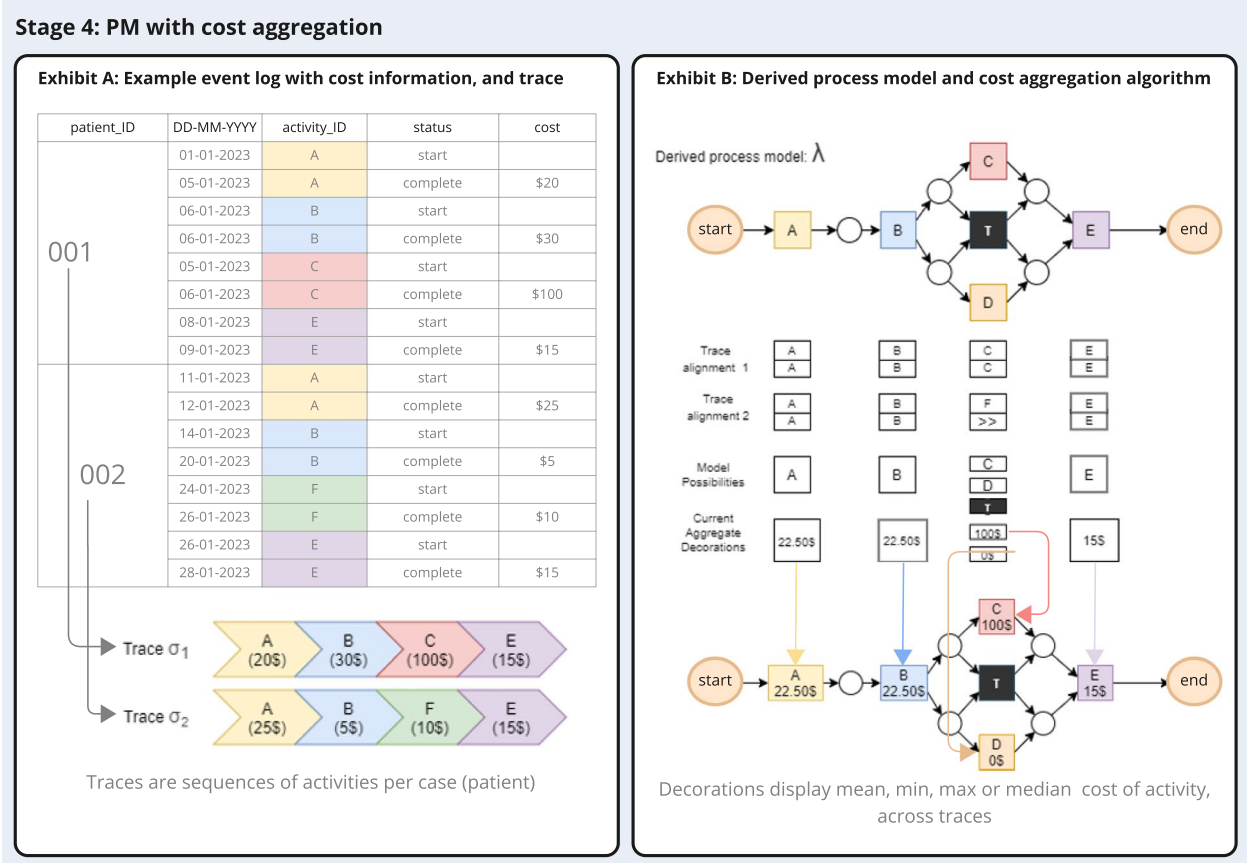


Fig. 4 Explanatory diagram depicting how the aggregation algorithm uses the data provided in the event log (exhibit A), transforms it into traces with cost information, and then derives cost statistics by aligning traces to compute mean, median, minimum, or maximum costs (exhibit B)

technique can account for the temporal nature of care, as the costs of e.g. receiving chemotherapy during late-stage cancer are higher than early-stage. In future, if protocol changes are introduced to e.g. circumvent the use of Mfolfox 6 during stage C CRC, the cost and duration impact of this change can be traced using cost mining.

Discussion

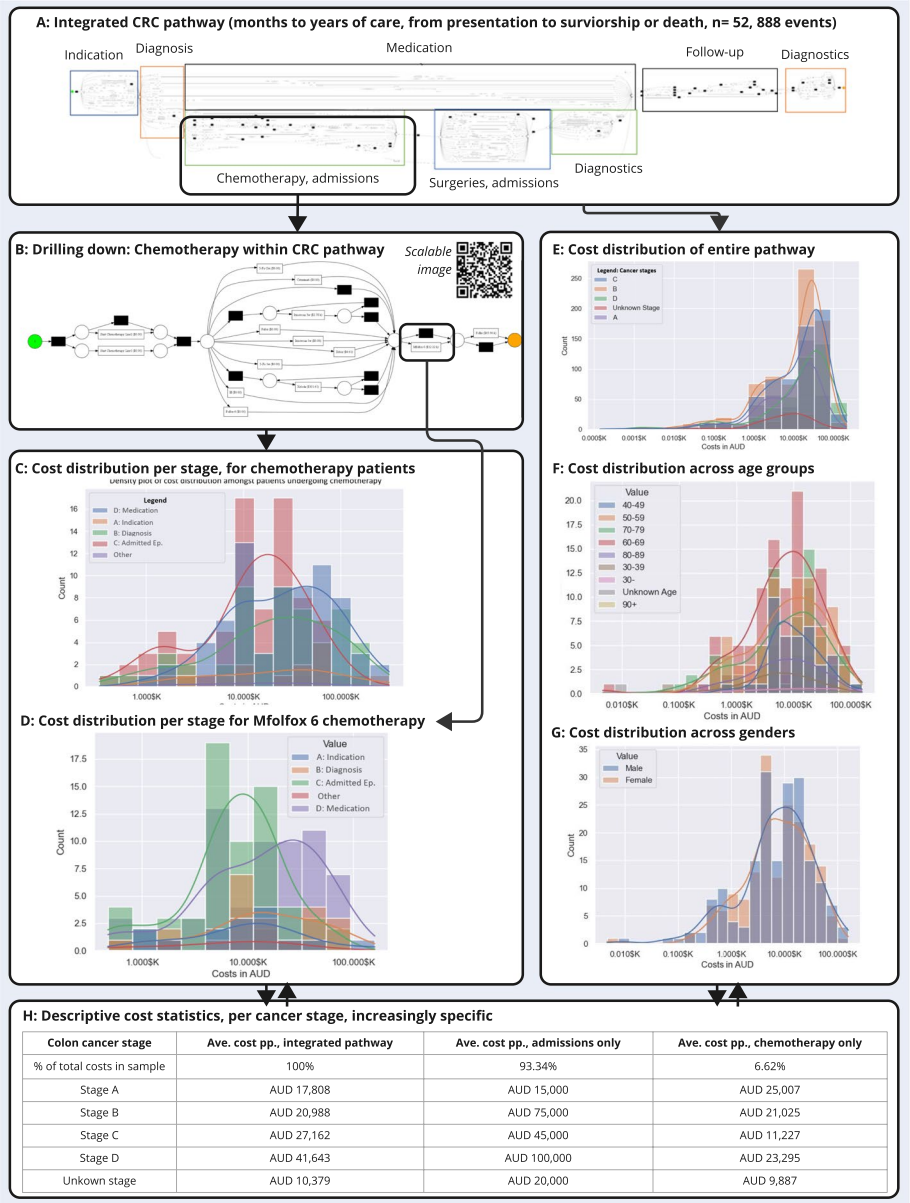
In this methodological paper, we draw on recent PM work in healthcare settings [13, 18, 25, 31, 41, 46] to develop and trial a method to support VBHC. Because cost mining aggregates cost information across entire patient journeys using real life data, this method translates large volumes of data into useful and practical information with which care can be made more efficient, accessible, and sustainable. In doing so, we have answered several recent calls for research [47–50] and built on recent methodological work calling for PM with financial KPIs [30].

Applications for cost mining

This method is relevant to achieving process efficiency, cost reduction, improved resource allocation, continuous process improvement, and data driven medical decision-making to ensure financial sustainability in a landscape of increasing complexity.

At the international level, this method could facilitate financial benchmarking across different standards of care and healthcare systems by comparing large patient cohorts in terms of patient pathways, to identify high-cost or long-duration pathways to target with interventions. Thus, it would supplement ongoing analyses, or large retrospective or prospective cohort studies, by providing patient flow information alongside common health economic analyses [50].

At the national level, this method can aid researchers and policy-makers in tracing and evaluating increasing healthcare delivery variation, for instance in response to medical protocol changes over time, technological advancements in medicine, and digitalization of healthcare service delivery. This is particularly relevant in



Colon cancer stage	Ave. cost pp., integrated pathway	Ave. cost pp., admissions only	Ave. cost pp., chemotherapy only
% of total costs in sample	100%	93.34%	6.62%
Stage A	AUD 17,808	AUD 15,000	AUD 25,007
Stage B	AUD 20,988	AUD 75,000	AUD 21,025
Stage C	AUD 27,162	AUD 45,000	AUD 11,227
Stage D	AUD 41,643	AUD 100,000	AUD 23,295
Unkown stage	AUD 10,379	AUD 20,000	AUD 9,887

Fig. 5 Illustrative results gained from PM with cost aggregation for CRC pathway, with particular focus on chemotherapy, to show how the method supports 'drilling down' to understand where high costs are being incurred, for which patient groups, and which treatment modalities

countries that feature strong or increasing care concentration, such as the Netherlands [51]. Further, cost mining could uncover the long-term consequences of shifting standards of care, by mapping and aggregating the costs associated with specific procedural guidelines by comparing patient groups before and after policy changes, or across locations. Even in less fragmented systems (e.g., US) where patient-level data is more integrated, cost mining still holds relevance. Although one could directly determine costs from patient-level data, cost mining offers the ability to uncover underlying patterns, sequences, and relationships within the care process, which can complement traditional microcosting studies by providing contextual information, and by exploring how sequences or timing impact costs, outcomes, and durations.

At the clinical level, it can reveal whether specific patient groups are consuming disproportionately more care than others, as we have demonstrated in our CRC case, or face significantly longer or more invasive trajectories. This may also enable assessment of care equity by, for example, comparing advantaged to disadvantaged or

underrepresented patient groups. By exploring utilization patterns in a systematic way using cost mining, future research could identify whether disadvantaged groups are consuming more or less care than their counterparts, which opens up new avenues for prevention and intervention strategies relating to health equity. Moreover, this information would, in turn, provide valuable insights for future health technology assessments or cost-effectiveness assessments, enabling them to estimate the process and cost impact of e-health technologies from financial, sustainability, and equity perspectives [52]. Further, this method could be used to explore the economic impact of prevention, early diagnosis [21, 22, 53] and excessive routine diagnostics [54] or prescriptions [55] by assessing and comparing integrated pathways longitudinally.

Costs of CRC in Australia

The contribution of the present study is that we find that cancer stages relate to costs, and that costs of specific elements of CRC care are dependent on the relative timing in which they are administered during a patient's integrated pathway. Previous studies in New Zealand [56], England [57], the US [58], Europe [59], and Australia [21, 22], reported on costs of care for CRC cases in relation to control variables like age and sex. Building on this, we report treatment-specific factors like cancer stage as explanatory factors of cost variation. Only two prior studies found CRC costs to relate to cancer stage [22, 57]. Our results extend these findings by showing that stages B and C have the highest total costs, and stages C and D have the highest mean cost per patient, which suggests that treatment-related factors and timing influence costs. Whilst prior work focused on treatments [21, 58], we included primary care and life events and captured the integrated pathway, covering all treatments and events related to CRC. Importantly, our results show that chemotherapy costs depend on the cancer stage, with specific patient groups requiring high-cost regimens like Mfolfox 6 at specific stages (e.g., stage C) relating to high per-patient costs. These findings extend recent work and illustrate the benefits of mapping integrated patient pathways with data from multiple providers (e.g., GPs) to explore costs in relation to cancer stage and timing of treatments. By incorporating the entire pathway, we show that the total healthcare burden of CRC in Australia is predominantly related to inpatient episodes, but that per-patient costs within chemotherapy vary and relate to specific regimes in specific cancer stages. Future research should utilize cost mining to investigate whether preventative interventions or earlier screening and diagnosis lead to quicker patient pathways or comparatively lower-cost inpatient and chemotherapy episodes, given the significant correlation between cancer stage at the time of treatment and costs. Beyond CRC, future studies could expand on our algorithm to develop routine cost mining evaluations in other costly contexts, complementing and informing traditional economic and qualitative methods.

Limitations of cost mining

Cost mining has limitations inherent to PM and the use of historical patient data, namely significant data requirements, descriptive nature, and a lack of predictive power. The method primarily visualizes as-is states using retrospective data, describing costs faced by patients who have completed (parts of) their care trajectory. This may not reflect current costs for treatments with recent technological developments, and the analysis should be repeated periodically to discover new pathways as they occur.

Due to the descriptive nature of this analysis, the method requires significant volumes of data to be representative, and results must be interpreted cautiously. The method can uncover high-cost pathways and identify paths or patient groups that completed unusually costly pathways. However, the method cannot be used to judge whether medical decisions were cost-effective not, and the user must assume that pathways were chosen out of medical necessity. The resulting visualizations should therefore be used to uncover cost drivers to inform VBHC projects, or to identify patient groups that face unusually costly or lengthy treatments, and should be used in tandem with methods like micro costing or cost-effectiveness analyses [8], and qualitative approaches like realist evaluations that uncover situational or causal mechanisms [55]. Low patient numbers in specific branches of pathways are not problematic if the patient number is representative of the entire study population. Because the analysis is descriptive, it is sensitive to omissions, so excluded cost or activity data will result in an underestimation of cost statistics. Lastly, some contexts may be difficult to model with PM. Systems with free choice of GP and healthcare provider are challenging due to fragmented patient data across providers, necessitating manual linkage. In contrast, systems with seamless electronic health records, like those in the Netherlands, are easier to model as they capture all general and specialist care regardless of location.

Conclusion and future research

The cost mining method identified inpatient and chemotherapy episodes as particularly costly in Australian CRC care, driven by cancer stage, accounting for 99% of the $60.63 M AUD economic burden on the Australian health system (2012–2020). Our analysis underscores the benefits of linked registries and cost mining for assessing healthcare costs across integrated pathways to inform VBHC projects. Future research could extend

this method, and address some of its limitations, using predictive PM utilizing machine learning [60], to produce process maps that are not only actionable but also predictive. Additionally, our method relies on static cost estimates per activity using DRG data, whereas future work could develop algorithms that allow resource usage to vary per activity per patient, using cost equations [8].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-024-02446-5.

Supplementary Material 1.

Acknowledgements
The authors thank Dr Hui-li Wong, medical oncologist at the Peter MacCallum Cancer Centre and clinical research fellow at the Walter and Eliza Hall Institute, for her contribution.

Authors' contributions
SR developed the underlying method as part of a larger project under the supervision of FF, DD, MT, and MIJ. ML and FF reviewed the code. Results were reviewed, discussed and validated by content experts JE, PG, JM, VW, MJ, FF and MIJ. ML drafted the article with contributions from FF, SR and MIJ (introduction, literature review, methods and results sections). ML further drafted the discussion after presenting results to FF, JE, MT, PG, JM, VW, MJ, and MIJ. MJ and MIJ conceptualized the study, wrote the research proposal and obtained funding for the study. ML and SR developed the figures, and ML completed the revisions following feedback from SR, FF, DD, JE, MT, PG, JM, VW, MJ, FF and MIJ. The final manuscript was extensively reviewed and approved by all authors.

Funding
This study was funded by the Victorian Cancer Agency, Health services research project in cancer survivorship—HSR19001. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Data availability
The linked dataset that was analyzed and used for this study is available from BioGrid (https://www.biogrid.org.au) on a secured server subject to ethics approval. The data is not publicly available, to preserve privacy and anonymity.

Declarations

Ethics approval and consent to participate
The study received ethical approval by the Royal Melbourne Hospital Ethics Board through BioGrid application number 202003/8 (Australia).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, the Netherlands. [2]Industrial Engineering and Business Information Systems, University of Twente, Enschede, the Netherlands. [3]University of Melbourne Centre for Cancer Research, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Parkville, VIC, Australia. [4]Center for Healthcare Operations Improvement and Research (CHOIR), University of Twente, Enschede, The Netherlands. [5]Centre for Health Policy, Melbourne School of Public and Global Health, Faculty of Medicine, Dentistry and Health Sciences, Parkville, VIC, Australia. [6]Department of Cancer Research, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia. [7]Cancer Research Program, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, Australia. [8]RMIT University, Melbourne, VIC, Australia. [9]School of Psychology, Faculty of Health, Deakin University, Melbourne, VIC, Australia. [10]Department of Medical Oncology, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia.

Received: 29 November 2023 Accepted: 16 December 2024
Published online: 27 December 2024

References

1. Karikios DJ, Schofield D, Salkeld G, Mann KP, Trotman J, Stockler MR. Rising cost of anticancer drugs in Australia. Intern Med J. 2014;44:44.
2. Smith TJ, Hillner BE. Bending the cost curve in cancer care. N Engl J Med. 2011;364:2060–5.
3. Chen S, Cao Z, Prettner K, Kuhn M, Yang J, Jiao L, et al. Estimates and projections of the global economic cost of 29 cancers in 204 countries and territories from 2020 to 2050. JAMA Oncol. 2023. https://doi.org/10.1001/jamaoncol.2022.7826.
4. Leusder M, Porte P, Ahaus K, van Elten H. Cost measurement in value-based healthcare: a systematic review. BMJ Open. 2022;12: e066568.
5. Keel G, Muhammad R, Savage C, Spaak J, Gonzalez I, Lindgren P, et al. Time-driven activity-based costing for patients with multiple chronic conditions: a mixed-method study to cost care in a multidisciplinary and integrated care delivery centre at a university-affiliated tertiary teaching hospital in Stockholm, Sweden. BMJ Open. 2020;10:e032573.
6. Alves RJV, Etges APB da S, Neto GB, Polanczyk CA. Activity-based costing and time-driven activity-based costing for assessing the costs of cancer prevention, diagnosis, and treatment: a systematic review of the literature. Value Health Reg Issues. 2018;17:142–7.
7. Rafiq M, Keel G, Mazzocato P, Spaak J, Guttmann C, Lindgren P, et al. Extreme consumers of health care: patterns of care utilization in patients with multiple chronic conditions admitted to a novel integrated clinic. J Multidiscip Healthc. 2019;12:1075–83.
8. Leusder M, van Elten HJ, Ahaus K, Hilders CGJM, van Santbrink EJP. Protocol for improving the costs and outcomes of assistive reproductive technology fertility care pathways: a study using cost measurement and process mining. BMJ Open. 2023;13: e067792.
9. Llewellyn S, Begkos C, Ellwood S, Mellingwood C. Public value and pricing in English hospitals: value creation or value extraction? Crit Perspect Account. 2022;85: 102247.
10. Ederhof M, Ginsburg PB. "meaningful use" of cost-measurement systems - incentives for health care providers. N Engl J Med. 2019;381:4–6.
11. Tan SS, Serdén L, Geissler A, van Ineveld M, Redekop K, Heurgren M. DRGs and cost accounting: which is driving which? In: Busse R, Geissler A, Quentin W, editors. Diagnosis-related groups in Europe: moving towards transparency, efficiency and quality in hospitals. European 1574 Z. Špacírová et al. 1 3 observatory on health systems and policies series, Maidenhead, Open University Press McGraw-Hill, Berkshire (2011). 2011. p. 59–74.
12. Špacírová Z, Epstein D, Espín J. Are costs derived from diagnosis-related groups suitable for use in economic evaluations? A comparison across nine European countries in the European Healthcare and Social Cost Database. Eur J Health Econ. 2022;23:1563–75.
13. van Hulzen G, Martin N, Depaire B, Souverijns G. Supporting capacity management decisions in healthcare using data-driven process simulation. J Biomed Inform. 2022;129:104060.
14. Agostinelli S, Covino F, D'Agnese G, De Crea C, Leotta F, Marrella A. Supporting governance in healthcare through process mining: a case study. IEEE Access. 2020;8:186012–25.
15. Benevento E, Aloini D, Squicciarini N, Dulmin R, Mininno V. Queue-based features for dynamic waiting time prediction in emergency department. Meas Bus Excell. 2019;23:458–71.
16. Aguirre JA, Torres AC, Pescoran ME. Evaluation of operational process variables in healthcare using process mining and data visualization techniques. Health. 2019;7:19.
17. Canjels KF, Imkamp MSV, Boymans TA, Vanwersch RJB. Unraveling and improving the interorganizational arthrosis care process at Maastricht UMC+: an illustration of an innovative, combined application of data and

process mining. In: Business Process Management. Aachen: ceur-ws.org; 2019. p. 178–89. https://ceur-ws.org/Vol-2428/paper16.pdf.
18. Cho M, Song M, Park J, Yeom SR, Wang IJ, Choi BK. Process mining-supported emergency room process performance indicators. Int J Environ Res Public Health. 2020;17;6290. https://doi.org/10.3390/ijerph17176290.
19. Wicky A, Gatta R, Latifyan S, Micheli RD, Gerard C, Pradervand S, et al. Interactive process mining of cancer treatment sequences with melanoma real-world data. Front Oncol. 2023;13: 1043683.
20. Jacobs K, Marcon G, Witt D. Cost and performance information for doctors: an international comparison. Manag Acc Res. 2004;15:337–54.
21. Goldsbury DE, Yap S, Weber MF, Veerman L, Rankin N, Banks E, et al. Health services costs for cancer care in Australia: estimates from the 45 and up study. PLoS One. 2018;13: e0201552.
22. Goldsbury DE, Feletto E, Weber MF, Haywood P, Pearce A, Lew JB, et al. Health system costs and days in hospital for colorectal cancer patients in New South Wales, Australia. PLoS One. 2021;16:e0260088.
23. Atkins ER, Geelhoed EA, Knuiman M, Briffa TG. One third of hospital costs for atherothrombotic disease are attributable to readmissions: a linked data analysis. BMC Health Serv Res. 2014;14: 338.
24. van der Spoel S, van Keulen M, Amrit C. Process prediction in noisy data sets: a case study in a dutch hospital. In: Lecture notes in business information processing. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 60–83.
25. Phan R, Augusto V, Martin D, Sarazin M. Clinical pathway analysis using process mining and discrete-event simulation: an application to incisional hernia. In: 2019 Winter Simulation Conference (WSC). National Harbour, MD, USA. 2019. p. 1172–83. https://doi.org/10.1109/WSC40007.2019.9004944.
26. Gerhardt R, Valiati JF, Canto dos Santos JV. An investigation to identify factors that lead to delay in healthcare reimbursement process: a Brazilian case. Big Data Research. 2018;13:11–20.
27. Huang H, Jin T, Wang J. Extracting clinical-event-packages from billing data for clinical pathway mining. In: Xing C, Zhang Y, Liang Y, (eds) Smart Health. ICSH 2016. Lecture Notes in Computer Science, vol 10219. Cham: Springer International Publishing; 2016. p. 19–31. https://doi-org.eur.idm.oclc.org/10.1007/978-3-319-59858-1_3.
28. Augusto A, Deitz T, Faux N, Manski-Nankervis J-A, Capurro D. Process mining-driven analysis of COVID-19's impact on vaccination patterns. J Biomed Inform. 2022;130: 104081.
29. Vathy-Fogarassy Á, Vassányi I, Kósa I. Multi-level process mining methodology for exploring disease-specific care processes. J Biomed Inform. 2022;125: 103979.
30. De Roock E, Martin N. Process mining in healthcare - an updated perspective on the state of the art. J Biomed Inform. 2022;127: 103995.
31. Munoz-Gama J, Martin N, Fernandez-Llatas C, Johnson OA, Sepúlveda M, Helm E, et al. Process mining for healthcare: characteristics and challenges. J Biomed Inform. 2022;127:103994.
32. Litchfield I, Hoye C, Shukla D, Backman R, Turner A, Lee M, et al. Can process mining automatically describe care pathways of patients with long-term conditions in UK primary care? A study protocol. BMJ Open. 2018;8:e019947.
33. Quintano Neira RA, Hompes BFA, de Vries JGJ, Mazza BF, Simões de Almeida SL, Stretton E, et al. Analysis and optimization of a sepsis clinical pathway using process mining. In: Business process management workshops. Cham: Springer International Publishing; 2019. p. 459–70.
34. Balakhontceva MA, Funkner AA, Semakova AA, Metsker OG, Zvartau NE, Yakovlev AN, et al. Holistic modeling of chronic diseases for recommendation elaboration and decision making. Procedia Comput Sci. 2018;138:228–37.
35. Huang Z, Dong W, Ji L, Yin L, Duan H. On local anomaly detection and analysis for clinical pathways. Artif Intell Med. 2015;65:167–77.
36. Poelmans J, Dedene G, Verheyden G, Mussele V der, Viaene S, Peters E. Combining business process and data discovery techniques for analyzing and improving integrated care pathways. In: Advances in data mining applications and theoretical aspects. 2010. p. 505–17.
37. Toth K, Machalik K, Fogarassy G, Vathy-Fogarassy A. Applicability of process mining in the exploration of healthcare sequences. In: 2017 IEEE 30th Neumann Colloquium (NC). Budapest: IEEE; 2017. p. 151–6.
38. Marazza F, Bukhsh FA, Geerdink J, Vijlbrief O, Pathak S, van Keulen M, et al. Automatic process comparison for subpopulations: application in cancer care. Int J Environ Res Public Health. 2020;17: 5707.
39. Ibanez-Sanchez G, Fernandez-Llatas C, Martinez-Millana A, Celda A, Mandingorra J, Aparici-Tortajada L, et al. Toward value-based healthcare through interactive process mining in emergency rooms: the stroke case. Int J Environ Res Public Health. 2019;16:1783.
40. van der Aalst W. Data science in action. In: van der Aalst W, editor. Process mining: data science in action. Berlin, Heidelberg: Springer Berlin Heidelberg; 2016. p. 3–23.
41. Andrews R, Goel K, Corry P, Burdett R, Wynn MT, Callow D. Process data analytics for hospital case-mix planning. J Biomed Inform. 2022;129: 104056.
42. Saint J, Fan Y, Singh S, Gasevic D, Pardo A. Using process mining to analyse self-regulated learning: a systematic analysis of four algorithms. In: LAK21: 11th International Learning Analytics and Knowledge Conference. New York: ACM; 2021. p. 333–43.
43. Maldonado-Mahauad J, Pérez-Sanagustín M, Kizilcec RF, Morales N, Munoz-Gama J. Mining theory-based patterns from big data: identifying self-regulated learning strategies in massive open online courses. Comput Human Behav. 2018;80:179–96.
44. Malmberg J, Järvelä S, Järvenoja H, Panadero E. Promoting socially shared regulation of learning in CSCL: Progress of socially shared regulation among high- and low-performing groups. Comput Human Behav. 2015;52:562–72.
45. Berti A, van der Aalst WMP. A novel token-based replay technique to speed up conformance checking and process enhancement. In: Koutny M, Kordon F, Pomello L, editors. Transactions on petri nets and other models of concurrency XV. Berlin, Heidelberg: Springer Berlin Heidelberg; 2021. p. 1–26.
46. Lim J, Kim K, Song M, Yoo S, Baek H, Kim S, et al. Assessment of the feasibility of developing a clinical pathway using a clinical order log. J Biomed Inform. 2022;128: 104038.
47. Born KB, Levinson W, Vaux E. Choosing Wisely and the climate crisis: a role for clinicians. BMJ Qual Saf. 2023. https://doi.org/10.1136/bmjqs-2023-015928.
48. Zimmerman JJ, Harmon LA, Smithburger PL, Chaykosky D, Heffner AC, Hravnak M, et al. Choosing wisely for critical care: the next five. Crit Care Med. 2021;49:472–81.
49. Robert G, Sarre S, Maben J, Griffiths P, Chable R. Exploring the sustainability of quality improvement interventions in healthcare organisations: a multiple methods study of the 10-year impact of the "productive ward: releasing time to care" programme in English acute hospitals. BMJ Qual Saf. 2020;29:31–40.
50. Martin N, De Weerdt J, Fernández-Llatas C, Gal A, Gatta R, Ibáñez G, et al. Recommendations for enhancing the usability and understandability of process mining in healthcare. Artif Intell Med. 2020;109:101962.
51. Gajadien CS, Dohmen PJG, Eijkenaar F, Schut FT, van Raaij EM, Heijink R. Financial risk allocation and provider incentives in hospital-insurer contracts in The Netherlands. Eur J Health Econ. 2023;24:125–38.
52. Granath A, Eriksson K, Wikström L. Healthcare workers' perceptions of how eHealth applications can support self-care for patients undergoing planned major surgery. BMC Health Serv Res. 2022;22:844.
53. McGarvey N, Gitlin M, Fadli E, Chung KC. Increased healthcare costs by later stage cancer diagnosis. BMC Health Serv Res. 2022;22:1155.
54. Moriates C. How can we finally reduce repetitive routine laboratory tests for hospitalised patients? BMJ Qual Saf. 2023. https://doi.org/10.1136/bmjqs-2023-016315.
55. Luetsch K, Wong G, Rowett D. A realist synthesis of educational outreach visiting and integrated academic detailing to influence prescribing in ambulatory care: why relationships and dialogue matter. BMJ Qual Saf. 2023. https://doi.org/10.1136/bmjqs-2022-015498.
56. Blakely T, Atkinson J, Kvizhinadze G, Wilson N, Davies A, Clarke P. Patterns of cancer care costs in a country with detailed individual data. Med Care. 2015;53:302–9.
57. Laudicella M, Walsh B, Burns E, Smith PC. Cost of care for cancer patients in England: evidence from population-based patient-level data. Br J Cancer. 2016;114:1286–92.
58. Mariotto AB, Robin Yabroff K, Shao Y, Feuer EJ, Brown ML. Projections of the cost of cancer care in the United States: 2010–2020. J Natl Cancer Inst. 2011;103:117–28.

59. Henderson RH, French D, Maughan T, Adams R, Allemani C, Minicozzi P, et al. The economic burden of colorectal cancer across Europe: a population-based cost-of-illness study. Lancet Gastroenterol Hepatol. 2021;6:709–22.
60. Pishgar M, Harford S, Theis J, Galanter W, Rodríguez-Fernández JM, Chaisson LH, et al. A process mining- deep learning approach to predict survival in a cohort of hospitalized COVID-19 patients. BMC Med Inform Decis Mak. 2022;22:194.

Publisher's Note