Process mining on actual treatment patterns

Lionel Perrier[1] et al.

Joint work with Omar Rifki[2], Zhihao Peng[3], Xiaolan Xie[3],

Int. J. Production Research, https://doi.org/10.1080/00207543.2024.2427888

[1]Centre Léon Bérard, CNRS, Université Lumière Lyon 2, Université Jean Monnet Saint-Etienne, Emlyon Business School, GATE, France

[2]LISIC, Univ. Littoral Côte d'Opale, Calais, France.

[3]Centre Ingénierie et Santé, CNRS UMR6158 LIMOS, Mines Saint-Etienne

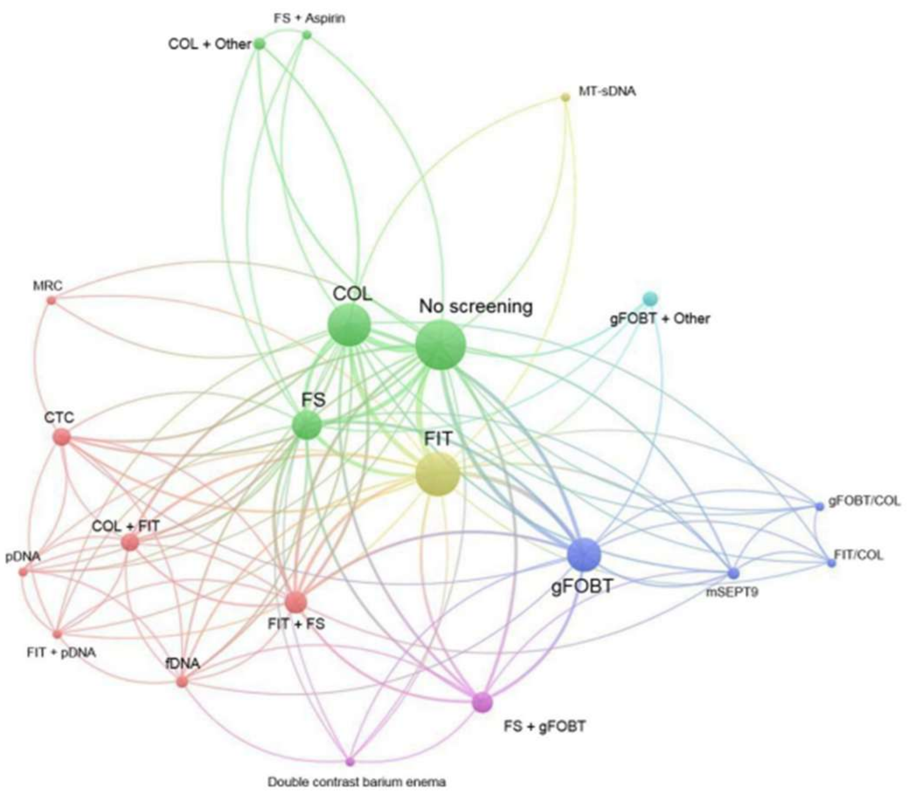
lionel.perrier@lyon.unicancer.fr

Content

1. State of art
2. New data-aware process mining
3. Discussion / limitations

1. State of art

- Data and process mining in healthcare are very interesting and important now, since data mining is widely used for classification and clustering *(refer to the publication of Yoo et al. 2012)*
- Innovative methodological frameworks are being developed and novel visualization schemes implemented *(refer to Giannoula et al. 2024)*
- There are also very interesting literature reviews in this field



FIT: fecal immunochemical test
COL: colonoscopy,
gFOBT: guaiac fecal occult blood test
FO: flexible sigmoidoscopy

Source : Mendivil J, Appierto M, Aceituno S, Comas M, Rué M. Economic evaluations of screening strategies for the early detection of colorectal cancer in the average-risk population: A systematic literature review. PLoS One. 2019 Dec 31;14(12):e0227251. doi: 10.1371/journal.pone.0227251. PMID: 31891647; PMCID: PMC6938313

References:

Chen K, Abtahi F, Carrero JJ, Fernandez-Llatas C, Seoane F. Process mining and data mining applications in the domain of chronic diseases: A systematic review. Artif Intell Med. 2023;144:102645. doi:10.1016/j.artmed.2023.102645

Giannoula A, Comas M, Castells X, et al. Exploring long-term breast cancer survivors’ care trajectories using dynamic time warping-based unsupervised clustering. J Am Med Inform Assoc. 2024;31(4):820-831. doi:10.1093/jamia/ocad251

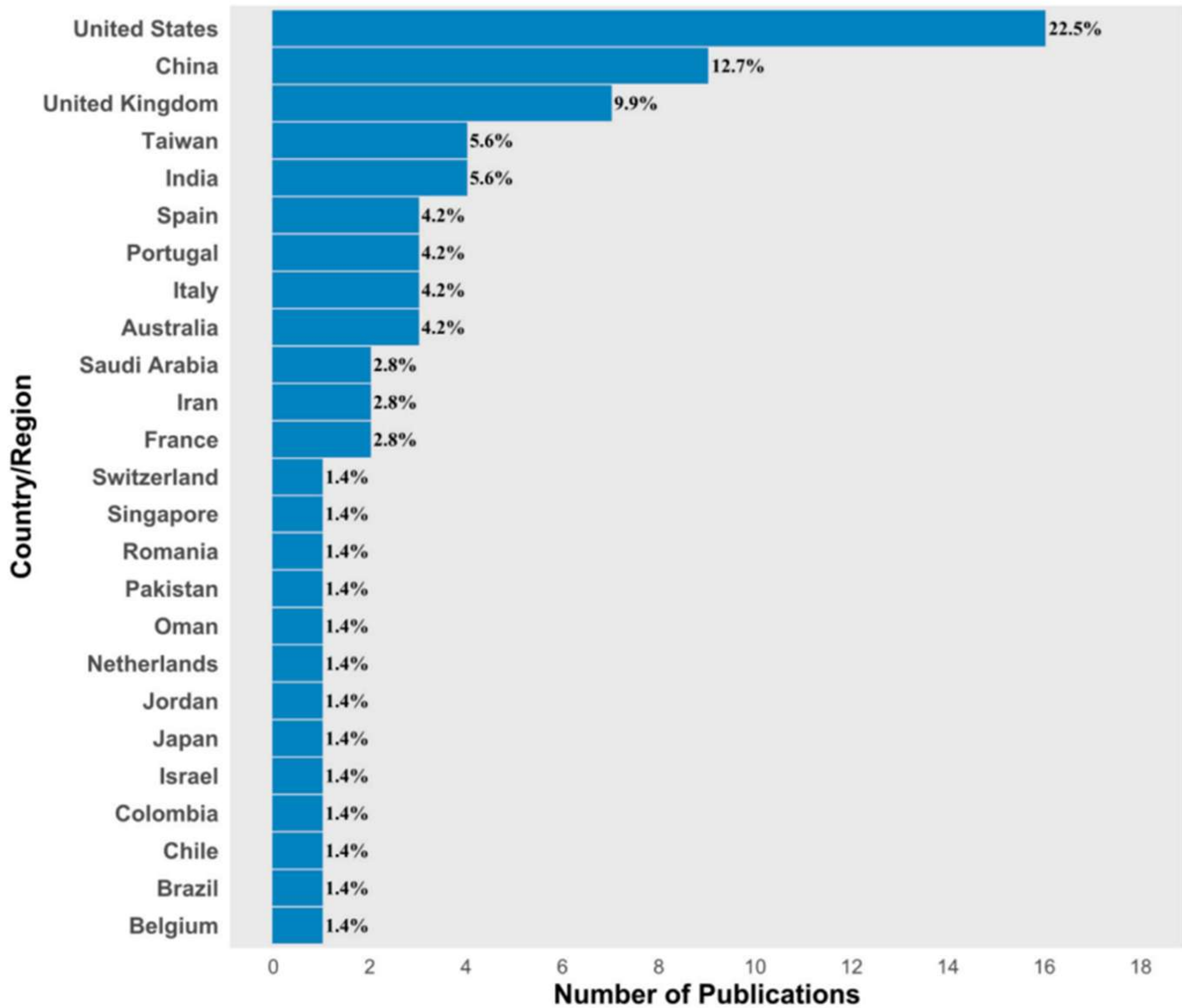
Kurniati A, Johnson O, Hogg D, Hall G. Process Mining in Oncology: a Literature Review. In: ; 2016. doi:10.1109/INFOCOMAN.2016.7784260

Kusuma GP, Kurniati AP, Rojas E, McInerney CD, Gale CP, Johnson OA. Process Mining of Disease Trajectories: A Literature Review. Stud Health Technol Inform. 2021;281:457-461. doi:10.3233/SHTI210200

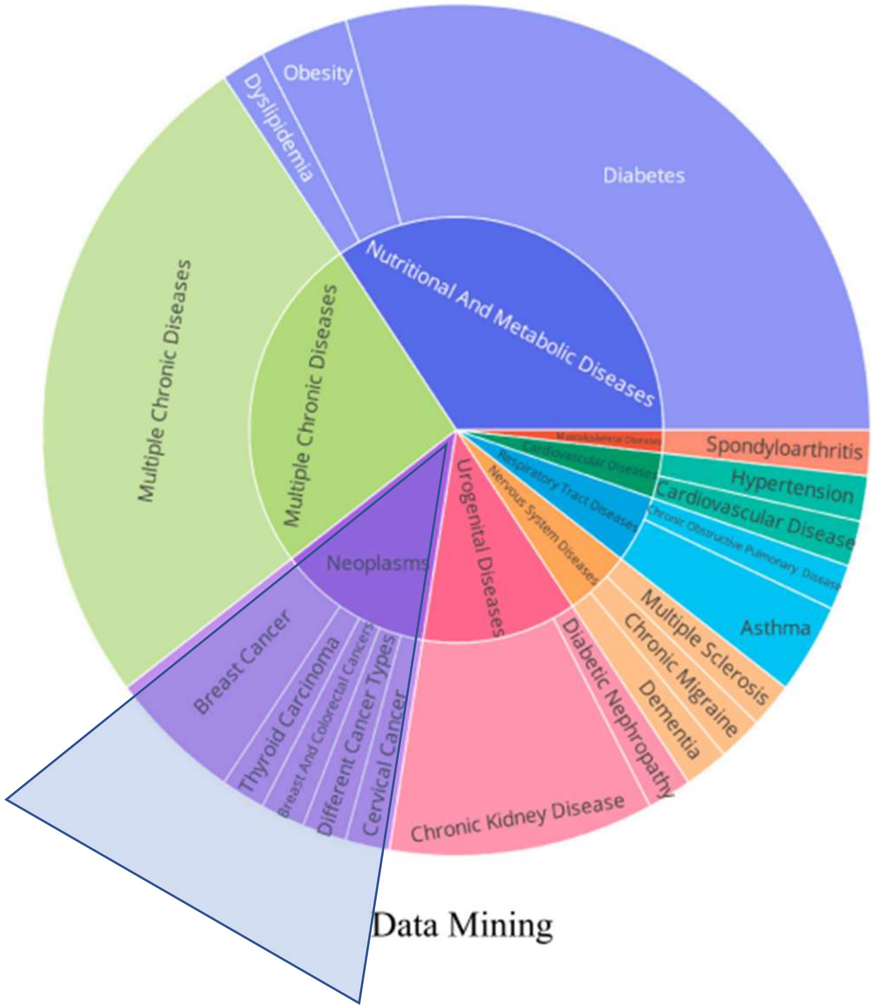
Mendivil J, Appierto M, Aceituno S, Comas M, Rué M. Economic evaluations of screening strategies for the early detection of colorectal cancer in the average-risk population: A systematic literature review. PLoS One. 2019 Dec 31;14(12):e0227251

Yoo I, Alafaireet P, Marinov M, et al. Data mining in healthcare and biomedicine: a survey of the literature. J Med Syst. 2012;36(4):2431-2448. doi:10.1007/s10916-011-9710-5

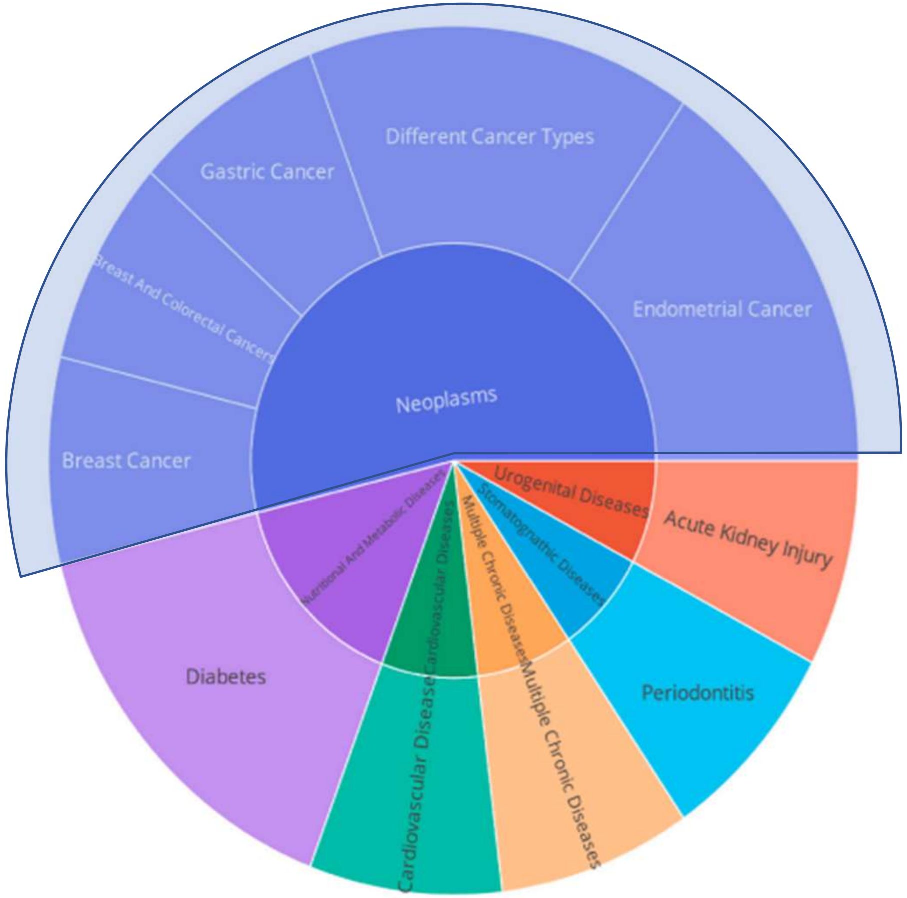
- **Chen's** on data and process mining in the field of chronic pathology found 71 articles published between 2000 and 2022
  - Only 13 of which (18%) apply process mining techniques
  - Articles were mainly published in the United States (22.5%)
  - Chen's review also shows that process mining is mostly used in oncology because it enables clinicians to analyze and detect complex healthcare processes and enhance cancer treatment



Source: Chen K, Abtahi F, Carrero JJ, Fernandez-Llatas C, Seoane F. Process mining and data mining applications in the domain of chronic diseases: A systematic review. Artif Intell Med. 2023;144:102645. doi:10.1016/j.artmed.2023.102645



Data Mining



Process Mining

Source: Chen K, Abtahi F, Carrero JJ, Fernandez-Llatas C, Seoane F. Process mining and data mining applications in the domain of chronic diseases: A systematic review. Artif Intell Med. 2023;144:102645. doi:10.1016/j.artmed.2023.102645

- **Kurniati et al. 2016**, identified 37 publications applying process mining techniques in oncology

- Gynecological cancers accounted for 24 of those articles, followed by
- breast cancer (four)
- colon, gastric, and lung cancer (three articles each)
- rectal cancer (two)
- and bladder, cervical, head and neck, and skin cancer (one article each)

Process Mining in Oncology: a Literature Review

Angelina Prima Kurniati[1,2,*], Owen Johnson[1], David Hogg[1], Geoff Hall[3,4]
[1] School of Computing, University of Leeds, Leeds, UK
[2] School of Computing, Telkom University, Bandung, Indonesia
[3] Leeds Institute of Cancer and Pathology, St James's University Hospital, UK
[4] School of Medicine, University of Leeds, Leeds, UK
*Corresponding author, email: scapk@leeds.ac.uk

References:

Kurniati A, Johnson O, Hogg D, Hall G. Process Mining in Oncology: a Literature Review. In: ; 2016. doi:10.1109/INFOCOMAN.2016.7784260

Kusuma GP, Kurniati AP, Rojas E, McInerney CD, Gale CP, Johnson OA. Process Mining of Disease Trajectories: A Literature Review. Stud Health Technol Inform. 2021;281:457-461. doi:10.3233/SHTI210200

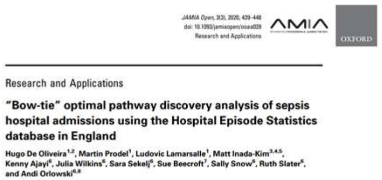
- **Kusuma et al.’s** review of studies that use process mining in identifying disease trajectories is also worth noting in this context

- Only four papers published to date have directly applied process mining to disease trajectory modelling
- There is currently very little research into the use of process mining for identifying disease trajectories, and highlighted a lack of awareness of these methods

#	Authors	Country/ Region	Data source	N data	Standard coding	Disease	PM Methodology	Model visualisation	Discovery algorithm	Trajectory approach	Conformance checking
1	Kusuma et al.[9]	Boston, USA	BIDMC Hospital	46520	ICD-9	General	PM^2	Directly-followed graph	iDHM	correlation measurement, binomial test	Replay fitness, precision, generalisation, k-folds cross validation
2	de Toledo et al.[10]	Spain	MBDS by the public healthcare provider	225,000	ICD-9	Type 2 Diabetes	KDD	Heuristics net	Heuristics miner and Fuzzy miner	n-grams	N/A
3	De Oliveira et al.[11]	England	NHS Hospital Episode Statistic	76,523	ICD-10	Sepsis	N/A	Private company app	Metaheuristics optimization algorithm	Metaheuristics optimisation algorithm	Replay fitness
4	Kusuma et al.[12]	N/A	(synthetic)	50	ICD-10	General	PM^2	Disco	iDHM	N/A	Replay fitness, precision, generalisation

Source: Kusuma GP, Kurniati AP, Rojas E, McInerney CD, Gale CP, Johnson OA. Process Mining of Disease Trajectories: A Literature Review. Stud Health Technol Inform. 2021;281:457-461. doi:10.3233/SHTI210200

- **De Oliveira et al. 2020**
  - Hospital Episode Statistics (HES) database for all patients in England with at least one hospital episode for sepsis present in any diagnosis within an episode spell between January 1, and December 31, 2016
  - Metaheuristic algorithm was used to create a “pathway” or process discovery model that best described the sequence of clinical events prior to and following the index hospitalization for sepsis
  - Label inputs: hospital episodes/hospital stays, event log using ICD-10 codes
  - A diagnosis of cancer, gastrointestinal disorders, pneumonia, or urinary tract infections most often directly preceded the hospitalization for sepsis



Source: Hugo De Oliveira, Martin Prodel, Ludovic Lamarsalle, Matt Inada-Kim, Kenny Ajayi, Julia Wilkins, Sara Sekelj, Sue Beecroft, Sally Snow, Ruth Slater, Andi Orlowski, “Bow-tie” optimal pathway discovery analysis of sepsis hospital admissions using the Hospital Episode Statistics database in England, *JAMIA Open*, Volume 3, Issue 3, October 2020, Pages 439–448, https://doi.org/10.1093/jamiaopen/ooaa039

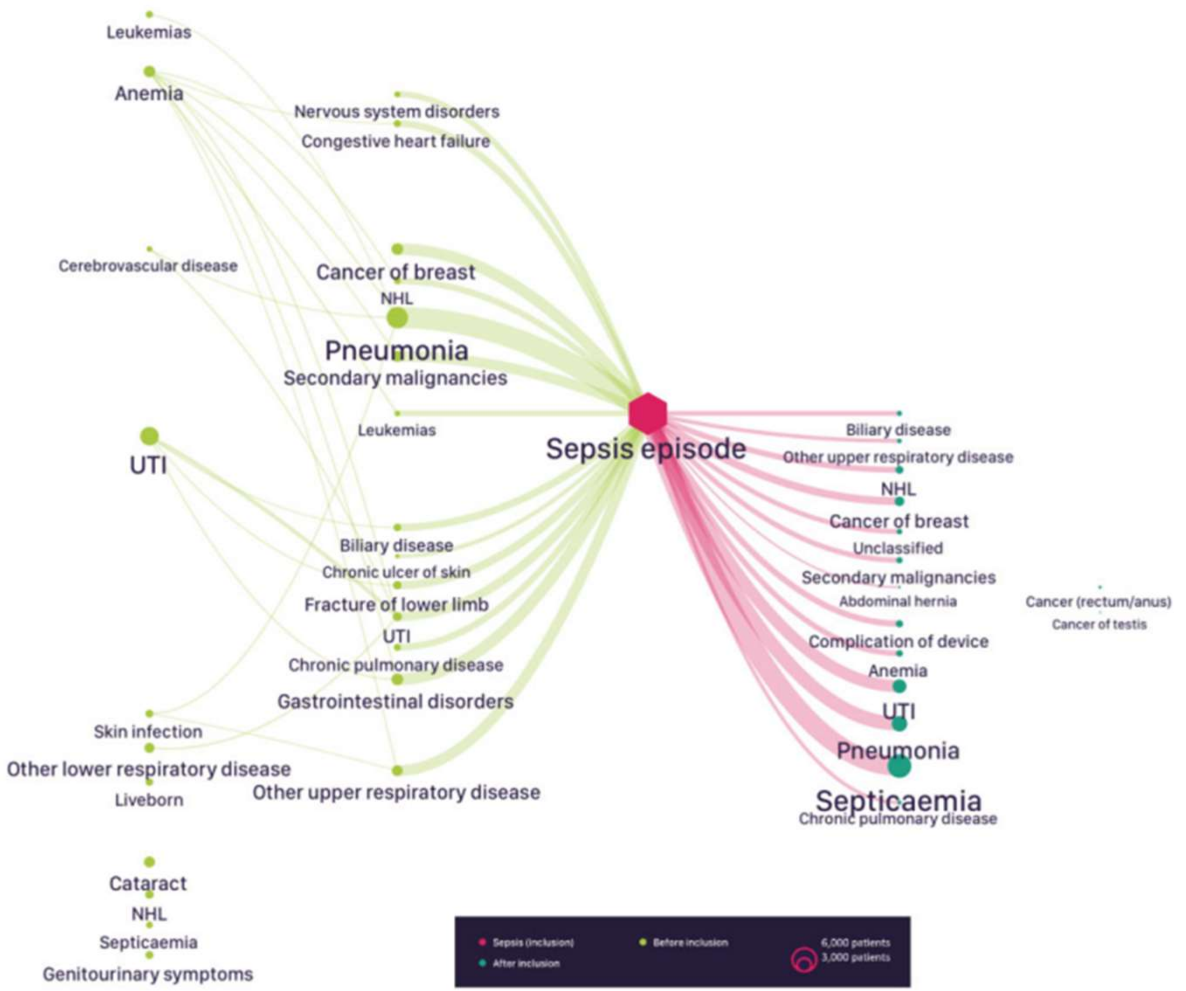
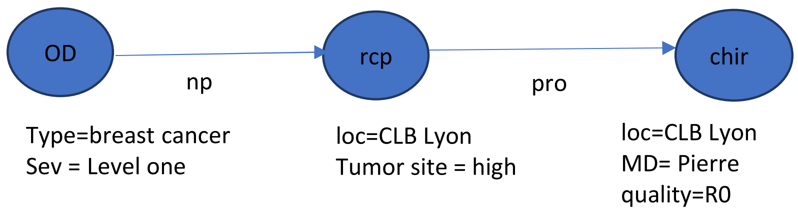


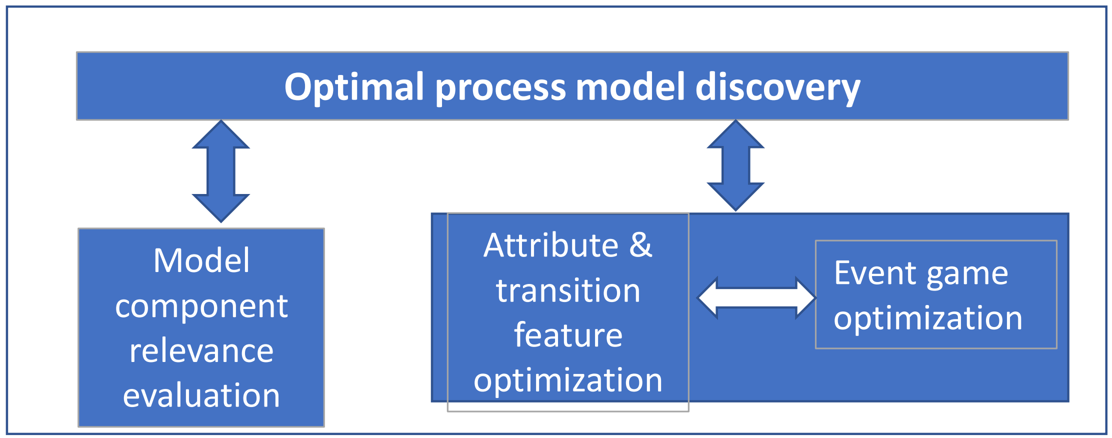
Figure 2. Bow-tie graph of the coded events in the 2 years before and 1 year after the index hospitalization for sepsis. The “bow-tie” graph is read from left to right, with circles representing event nodes of the process model (ie, coded events). The links (or edges) from each circle represent the time-ordered sequence of one coded event node following another. The sizes of nodes and links are proportional to the number of patients following this pathway. *Note*: The coded event “septicemia” contains a number of additional sepsis-related codes in addition to A40 or A41 (and their derivatives). See Supplementary Materials for full details of the HES ICD-10 codes included in this coded event.

2. New data-aware process mining

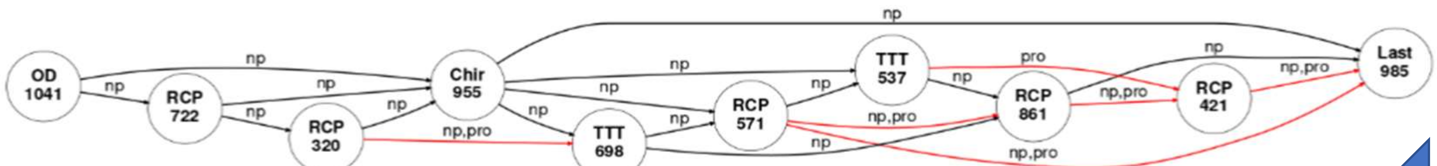
- ## Overall approach



1000+ care pathways with event attributes
&
transition features



The latent care pathway model

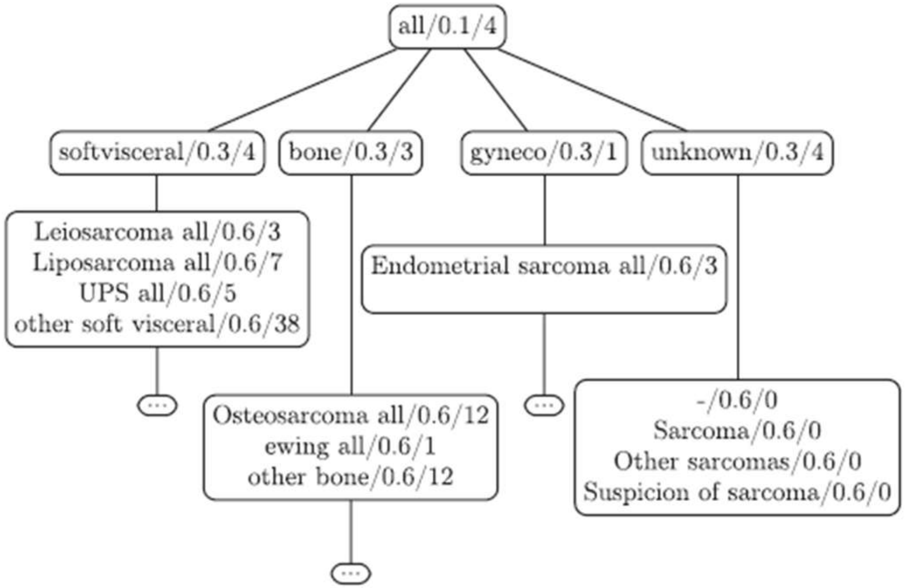


RCP1 722	RCP2 320	Chir 985	TTT1 698	RCP3 571	TTT2 537	RCP4 861	RCP5 421
soft vis 481	soft vis 203	soft vis 661	soft vis 491	soft vis 410	soft vis 401	soft vis 609	soft vis 314
loc in 722	loc in 320	adult 910	adult 662	loc in 570	adult 509	loc in 859	loc in 420
req in 442	req in 179	loc in 954	median 308	req in 334	large 291	req in 522	req in 256
			deep 381		deep 331		
			R0 386		R0 273		

Abbreviations: OD: original diagnosis; RCP: specialized MDTB; Chir: surgery; TTT: neoadjuvant/adjuvant treatment; Last: last contact; np: progression-free (black arcs); pro: disease progression (red arcs); Soft vis: Site of tumor category = Soft tissue or Viscera; R0: Quality of 1st surgery = R0 margin; Deep: depth of tumor = deep

- ## Input data modelling – Care event attributes

Value & macrovalue of attibutes (type)



Attribute	Root	Level-1
loc RCP	all/0.1/2	inside/0.3/71 outside/0.3/0
loc Chir	all/0.1/3	inside/0.3/32 outside/0.3/0 –/0.3/0
req centre	all/0.1/3	inside/0.3/44 outside/0.3/138 –/0.3/0
age	all/0.1/3	child/1/0 adult/1/0 –/1/0
depth	all/0.1/4	superficial/1/0 deep/1/0 deep+superficial/1/0 –/1/0
quality	all/0.1/4	R0/1/0 R1/1/0 R2/1/0 –/1/0
size	all/0.1/4	small/1/0 median/1/0 large/1/0 –/1/0
site	all/0.1/4	superior/0.3/16 inferior/0.3/14 trunk/0.3/15 others/0.3/45

Abbreviations: OD: original diagnosis; RCP: specialized MDTB; Chir: surgery; TTT: neoadjuvant/adjuvant treatment; Last: last contact; np: progression-free (black arcs); pro: disease progression (red arcs); Soft vis: Site of tumor category = Soft tissue or Viscera; R0: Quality of 1st surgery = R0 margin; Deep: depth of tumor = deep

- **Input data modelling - Cohort**

NETSARC all patients diagnosed with sarcoma in 2013 who had surgery for their primary tumor

2203 patients treated according to four care management strategies (1068, 108, 750, and 277 patients):

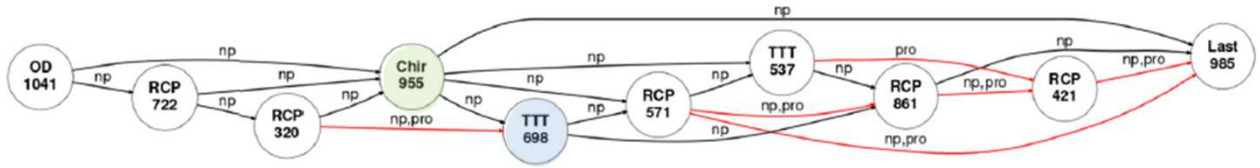
  - strategy 1 (1068) -complete initial management in the network with a Sarcoma MDTB before/after the initial surgery
  - strategy 2 (108)-outside initial management with a Sarcoma MDTB before the initial surgery
  - strategy 3 (750)-similar to 2 but with a Sarcoma MDTB after the initial surgery
  - strategy 4 (277)-outside initial management and no Sarcoma MDTB

Strategy 0 denotes all patients

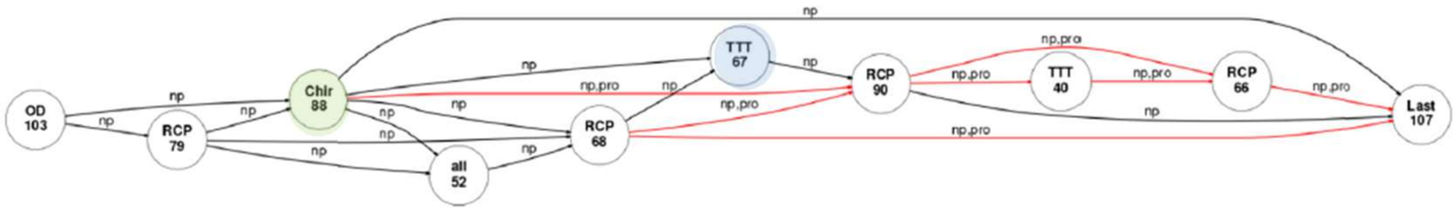
- **Some results – strategies 1-2-3**

Value of data-awareness:

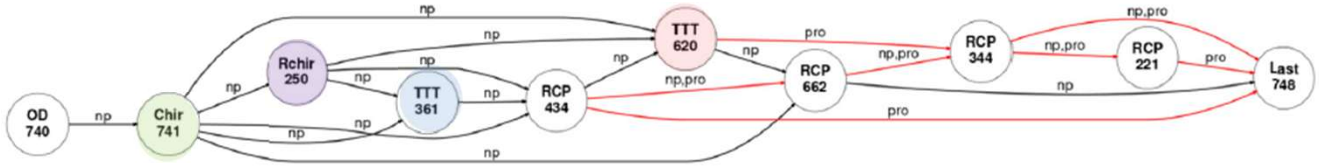
- Confirmation of strategies with surgery done inside in strategy 1 and outside in strategies 2 and 3
- TTT by far for deep tumor
- 2nd TTT in strategy 3 due to lower surgery quality R1
- RChir in strategy 3 (250/741) by far for lower quality R1 surgery



RCP1 722	RCP2 320	Chir 985	TTT1 698	RCP3 571	TTT2 537	RCP4 861	RCP5 421
soft vis 481 loc in 722 req in 442	soft vis 203 loc in 320 req in 179	soft vis 661 adult 910 loc in 954	soft vis 491 adult 662 median 308 deep 381 R0 386	soft vis 410 loc in 570 req in 334	soft vis 401 adult 509 large 291 deep 331 R0 273	soft vis 609 loc in 859 req in 522	soft vis 314 loc in 420 req in 256



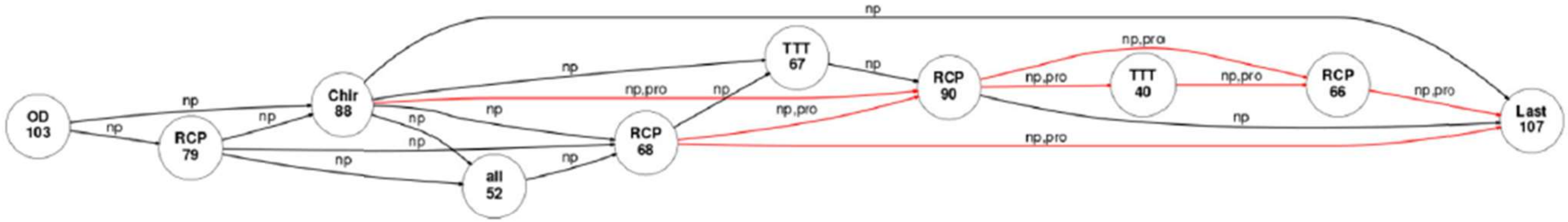
RCP1 79	Chir 88	all 52	RCP2 68	TTT1 67	RCP3 90	TTT2 40	RCP4 66
other soft 24 loc in 79 req in 50	other soft 31 adult 82 loc out 88	other soft 18 adult 36 large 11 deep 15 ...	other soft 23 loc in 68 req in 41	other soft 23 adult 62 median 27 deep 34 R0 32	other soft 31 loc in 90 req in 53	other soft 26 adult 39 median 19 deep 18 R0 15	other soft 23 loc in 66 req in 45



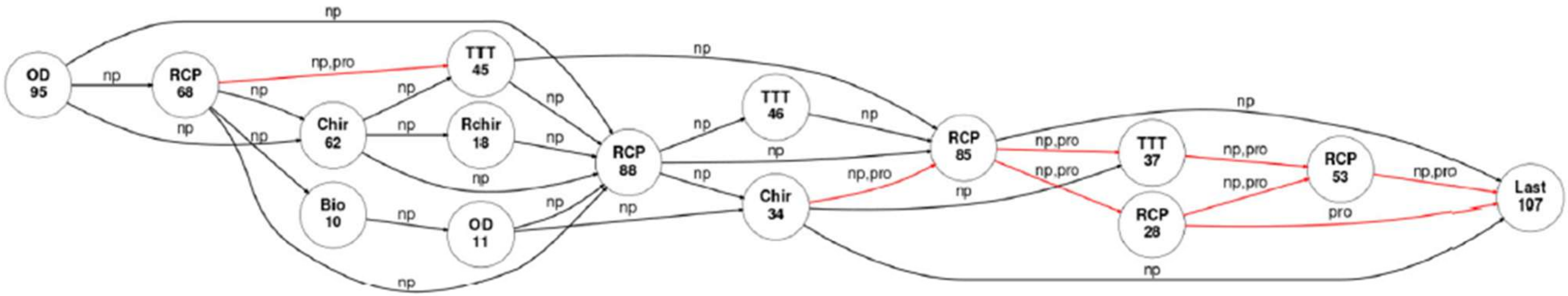
Chir 741	Rchir 250	TTT1 361	RCP1 434	TTT2 620	RCP2 662	RCP3 344	RCP4 221
soft vis 627 adult 734 loc out 741	soft vis 228 adult 249 R1 148	soft vis 313 adult 360 small 139 deep 195 R1 174	soft vis 373 loc in 434 req in 232	soft vis 519 adult 613 small 270 deep 324 R1 273	soft vis 558 loc in 662 req in 350	soft vis 305 loc in 344 req in 192	soft vis 200 loc in 218 req in 126

Abbreviations: OD: original diagnosis; RCP: specialized MDTB; Chir: surgery; TTT: neoadjuvant/adjuvant treatment; Last: last contact; np: progression-free (black arcs); pro: disease progression (red arcs); Soft vis: Site of tumor category = Soft tissue or Viscera; R0: Quality of 1st surgery = R0 margin; Deep: depth of tumor = deep

• Some results – higher precision (10 vs 15 nodes – strategy 2)



RCP1 79	Chir 88	all 52	RCP2 68	TTT1 67	RCP3 90	TTT2 40	RCP4 66
other soft 24	other soft 31	other soft 18	other soft 23	other soft 23	other soft 31	other soft 26	other soft 23
loc in 79	adult 82	adult 36	loc in 68	adult 62	loc in 90	adult 39	loc in 66
req in 50	loc out 88	large 11	req in 41	median 27	req in 53	median 19	req in 45
		deep 15		deep 34		deep 18	
		...		R0 32		R0 15	



RCP1 68	Chir1 62	Bio 10	TTT1 45	Rchir 18	RCP2 88	TTT2 46	Chir2 34	RCP3 85	TTT3 37	RCP4 28	RCP5 53
other soft 20	other soft 25	large 6	other soft 17	other soft 8	other soft 27	soft vis 30	Ewing 6	other soft 30	other soft 13	other soft 11	other soft 18
loc in 68	adult 59	deep 7	adult 41	adult 18	loc in 88	adult 44	adult 28	loc in 85	adult 35	loc in 28	loc in 53
req in 40	loc out 62	site Retrope-	median 19	R1 7	req in 55	median 19	loc out 34	req in 52	median 16	req in 19	req in 35
		ritoneum 4	deep 26			deep 24			deep 15		
		CHU Nancy 2	R0 20			R0 19			R0 14		

Abbreviations: OD: original diagnosis; RCP: specialized MDTB; Chir: surgery; TTT: neoadjuvant/adjuvant treatment; Last: last contact; np: progression-free (black arcs); pro: disease progression (red arcs); Soft vis: Site of tumor category = Soft tissue or Viscera; R0: Quality of 1st surgery = R0 margin; Deep: depth of tumor = deep

3. Discussion / limitations

- **Some interesting facts**

- Our new process mining[1] approach allows us to represent specialized MDTBs in the pathways of adult patients with soft tissue, visceral, or bone sarcoma

- There were significant differences between care strategies MDTB-labelled sarcoma before initial surgery and complete initial management in the network vs. MDTB-labelled sarcoma after initial surgery and initial management outside the network

- The event label "Second surgical excision/re-excision (Rchir)" and the attribute "R1 margin (histological positive margins)" both appear in care strategy MDTB-labelled sarcoma after initial surgery and initial management outside the network

- MDTB-labelled sarcoma appear later in the patient care pathways, when the proportion of patients whose disease has progressed increases

- These results are consistent with the medical literature

Reference:

[1] Rifki O, Peng Z, Perrier L, Xie X. Process mining with event attributes and transition features for care pathway modelling. International Journal of Production Research. 0(0):1-25. doi:10.1080/00207543.2024.2427888

- **Limitations**

  - It would be interesting to consider additional attributes in our process model, such as socioeconomic status, race/ethnicity, the distance between the patient’s home and the hospital, and insurance status, all of which may impact access to care

  - It would have been interesting to distinguish between radiotherapy and chemotherapy treatments; but this information is not available

  - The choice of attributes and event labels ultimately depends upon of the quality of the data

  - In France, this could be done by applying the process mining methods developed here to the database of the French NETSARC+ network matched to the French national health insurance database (SNDS). This could illustrate the costs involved in and value of strategic management approaches and provide useful insights to inform healthcare policymakers

Ethical statement: the study received approval by the French national commission for data privacy; Commission Nationale de l'Informatique et des Libertés (CNIL) the 21th November 2019 under the number 919360 (DR-2021-035)

Funding: this work was supported by a grant from the French Ministry of Health (PRME-18-0162). NetSARC+ is supported by the Institut National du Cancer (INCA) and DGOS. The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the report

Acknowledgments: the authors would like to thank the NetSarc network and community

Conflicts of Interest: the authors declare no conflict of interest

Thank you very much for your attention