School of Computer Science
FACULTY OF ENGINEERING

UNIVERSITY OF LEEDS

Process Mining in Healthcare Opportunities and Challenges

PRSB Professional Record Standards Body

BCS Faculty of Health and Care *and*
Professional Record Standards Body
20th March 2025
17.30-18.30

Owen Johnson

o.a.johnson@leeds.ac.uk

Associate Professor – Digital Health

Director of Impact, School of Computer Science

Co-Director, Leeds CDT for AI in Medical Diagnosis and Care

Programme Manager, MRes in Data Science for Healthcare (NHS England)

Process Mining in Healthcare – Opportunities and Challenges

UNIVERSITY OF LEEDS

An introduction to process mining in health and how it can help drive innovation in care, lower costs and improve patient outcomes

- The NHS is under pressure to improve its clinical and operational processes through better digital health and emerging AI.
- Process mining combines data science and process science methods for data-driven process improvement.
- The NHS is rich in data but using it effectively requires state-of-the-art tools, methods and skills.
- This talk will demonstrate how process mining can be used to improve NHS care pathways
- It will present the ClearPath method for process mining in healthcare developed at the University of Leeds
- … and discuss the opportunities and challenges in practice.

The Agenda
"Huge public rollout of AI"

UNIVERSITY OF LEEDS

Tuesday 13th January 2025

https://www.youtube.com/watch?v=o-AvkiRT9eA

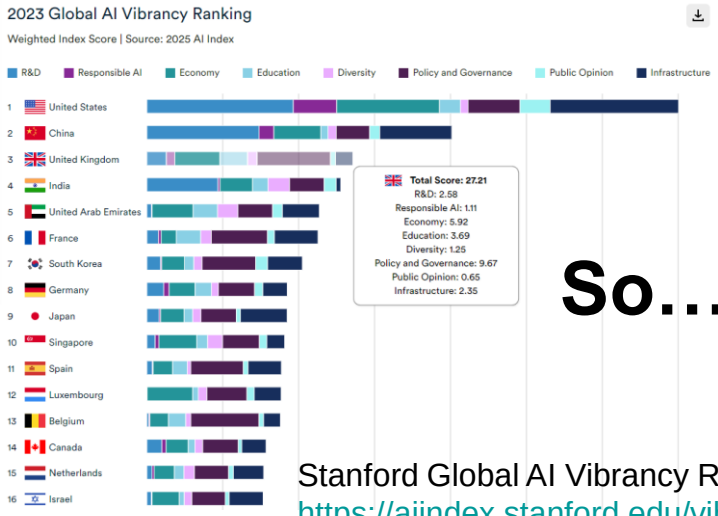
AI Opportunities Action Plan

The Government will:

- **Invest in the foundations of AI:** We need world-class computing and data infrastructure, access to talent and regulation (Section 1).
- Push hard on cross-economy AI adoption: The public sector should **rapidly pilot and scale AI** products and services and encourage the private sector to do the same. This will drive better experiences and outcomes for citizens and boost productivity (Section 2).
- **Position the UK to be an AI maker,** not an AI taker: As the technology becomes more powerful, we should be the best state partner to those building frontier AI. The UK should aim to have true national champions at critical layers of the AI stack so that the UK benefits economically from AI advancement and has influence on future AI's values, safety and governance (Section 3).

https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan

www.theguardian.com/politics/2025/jan/12/mainlined-into-uks-veins-labour-announces-huge-public-rollout-of-ai



Stanford Global AI Vibrancy Report Source: https://aiindex.stanford.edu/vibrancy/

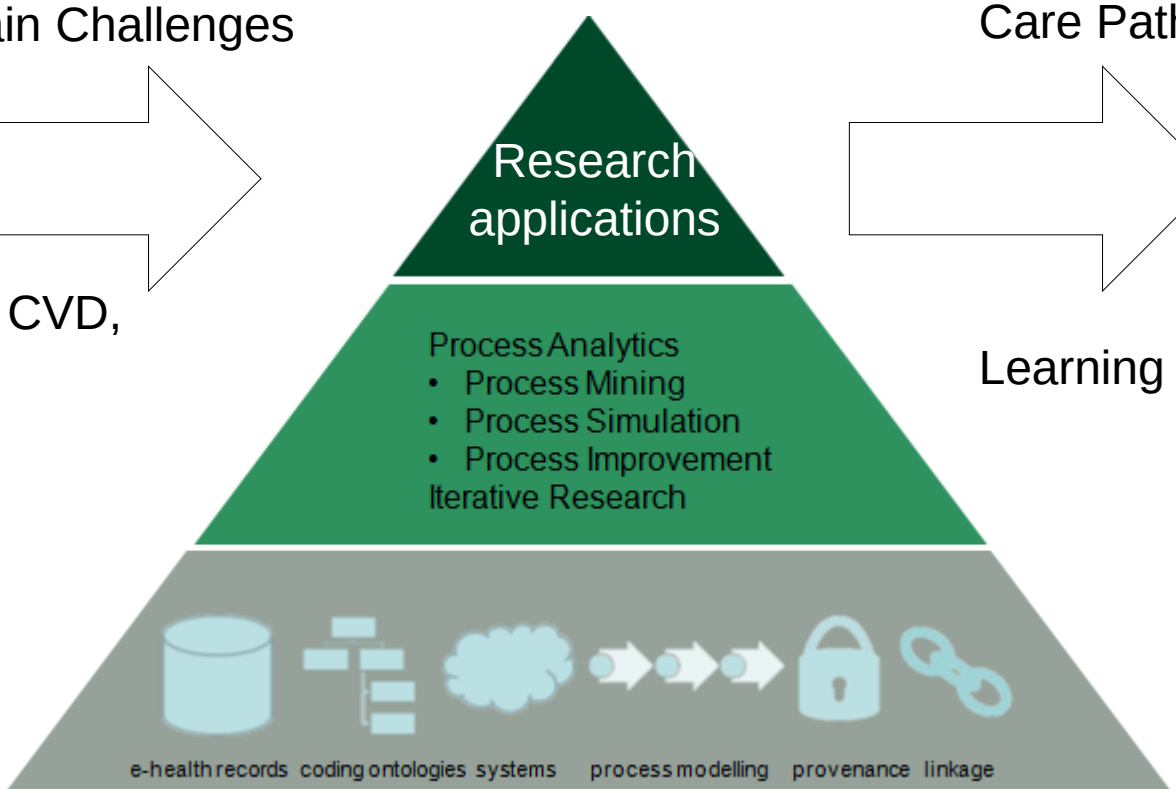
So… What do we do?

e-Health Records Research
School of Computer Science

UNIVERSITY OF LEEDS

Medical Domain Challenges

Cancer, MSK, CVD,
Diagnostics,
Emergency,
Dentistry,
etc.



Care Pathway Insights

Learning Health Systems


Owen


Angelina — Cancer
Sam C. — MSK
Guntur — Cardiovascular
Sam S — MSK
Amirah — Machine Learning
Frank — Dentistry
Nik — Frailty
Thamer — Machine Learning
Ciarán — Methods

Mission:
Design better care processes (with AI)

UNIVERSITY OF LEEDS

"Huge public rollout of AI"

Current → **Future**

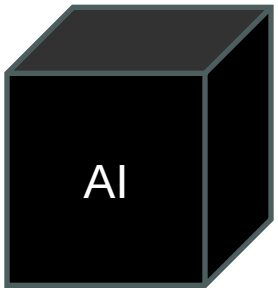
+

IMPLEMENTATION CHALLENGE
Where (in the process) will it be implemented?

What are the real-world implications?
How will the users be integrated?
What risks and bias?

DEVELOPMENT CHALLENGE
What AI?

What data will it use?
How was it trained?
What risks and bias?



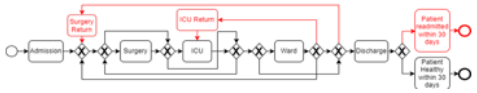
Images: ChatGPT 4.0

Research question:
How to design better care processes?

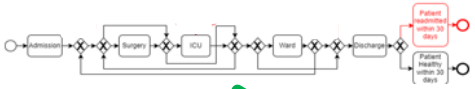
UNIVERSITY OF LEEDS

Current → **Future**

Analysis of the "As is" pathway



Design of the "Will be" pathway



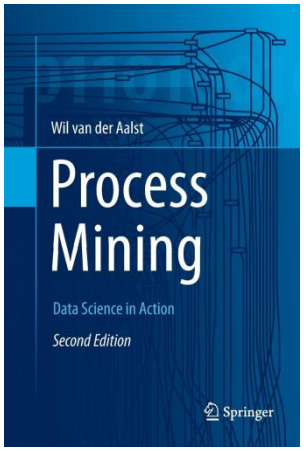
Pathway Improvement

Images: ChatGPT 4.0

Process Mining Since 2005

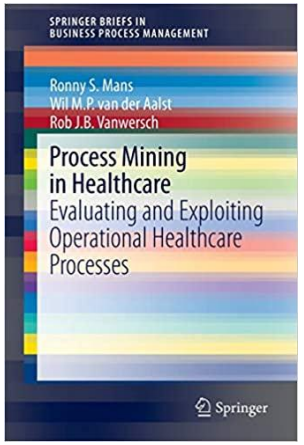
UNIVERSITY OF LEEDS

Process mining is to discover, monitor and improve real processes by extracting knowledge from event logs readily available in today's (information) systems.



Process mining includes:

- (automated) process discovery
- conformance checking (i.e., monitoring deviations by comparing model and log)
- Social network/ organizational mining
- automated construction of simulation models
- model extension
- model repair
- case prediction
- history-based recommendations



Unlike traditional approaches the goal is not to construct a single static model. Process mining techniques can be used to dynamically generate process maps based on the most recent data.

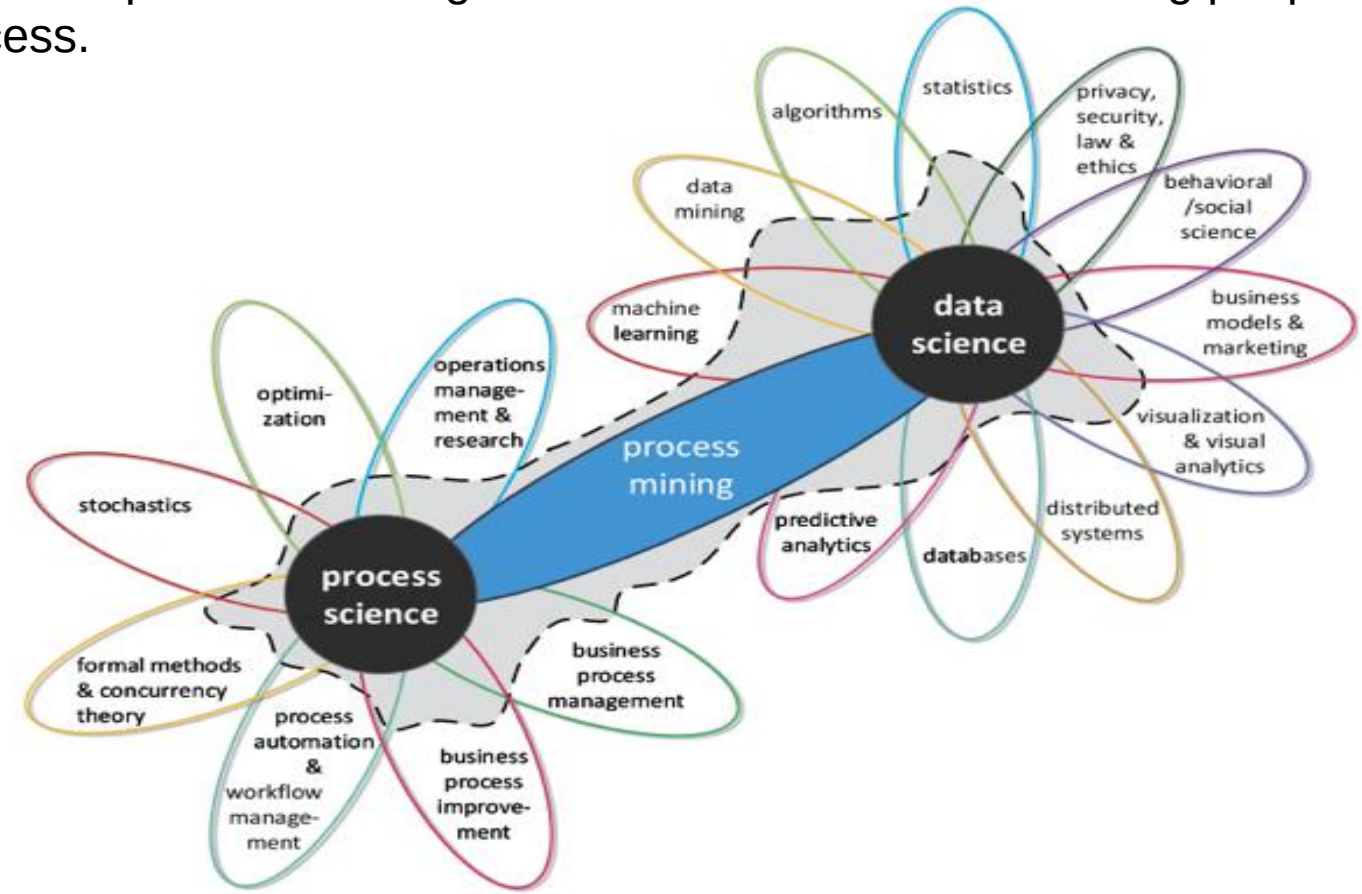
Professor Wil van der Aalst
(God father of process mining)
Department of Mathematics & Computer Science
Eindhoven University of Technology

http://www.win.tue.nl/ieeetfpm/lib/exe/fetch.php?media=shared:process_mining_manifesto-small.pdf

Process Mining
Is a Body of Knowledge

UNIVERSITY OF LEEDS

Process mining is a research discipline that bridges the gap between classical process model analysis and data science analysis. Process mining focuses on understanding real business processes using real data. In classical data mining people usually ignore the process.



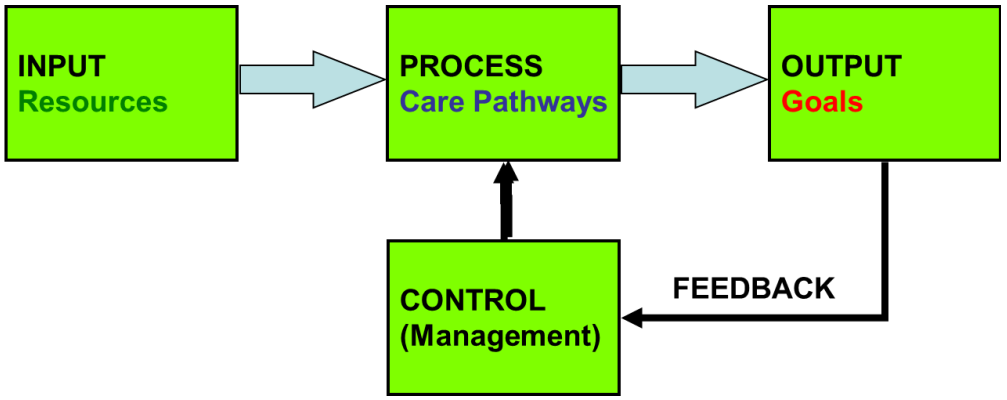
Van Eck, M.L., Lu, X., Leemans, S.J. and Van Der Aalst, W.M., 2015, June. PM^2: a process mining project methodology. In International Conference on Advanced Information Systems Engineering (pp. 297-313). Springer, Cham.

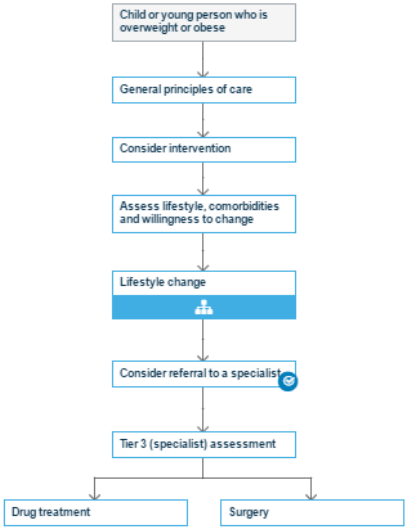
Process
Simple view

UNIVERSITY OF LEEDS

In health and social care, processes may be operational business processes (expenses, staff scheduling etc.) or care pathways…

Typical "Designed" pathway





Example: Managing children and young people who are overweight or obese
pathways.nice.org.uk

NB NICE withdrew their comprehensive set of care pathway guidelines in 2022

A Process Complex View

UNIVERSITY OF LEEDS

Typical "As is" pathway

Patients in poor health

Money £

- Resources
- Staff
- Assets

Patients in better health

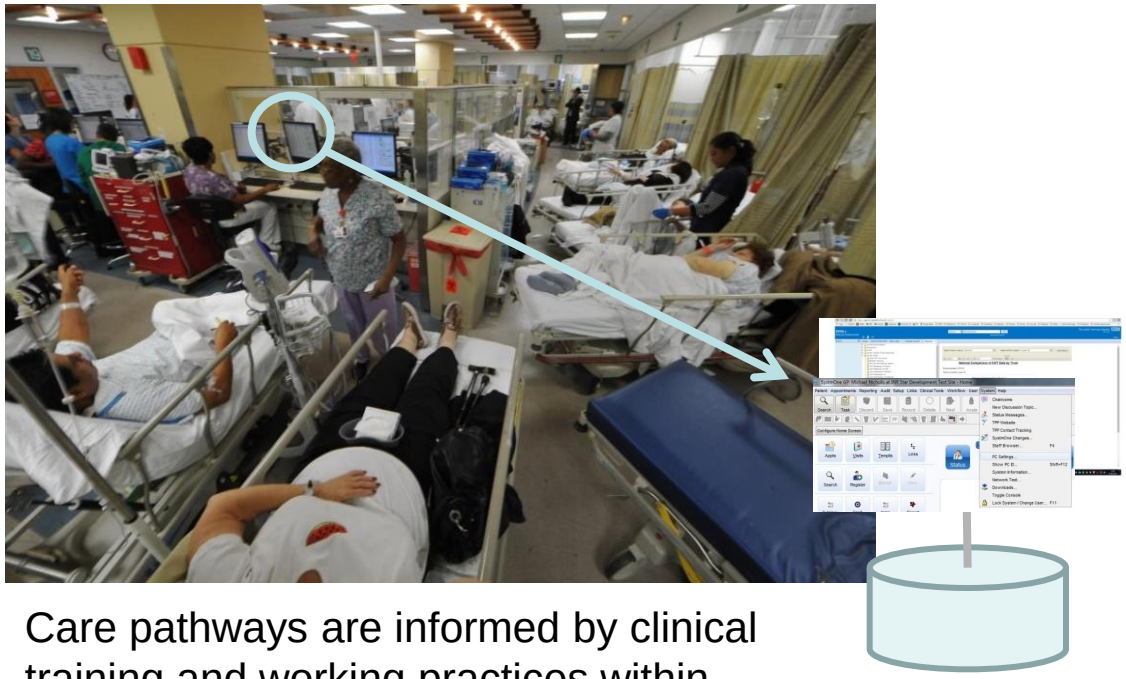
Quality Outcomes
- Waiting times
- Targets
- Customer satisfaction

Health Organisations are **complex systems**
They have non-linear behaviour and emergent properties

Business Processes in Healthcare
Care pathways

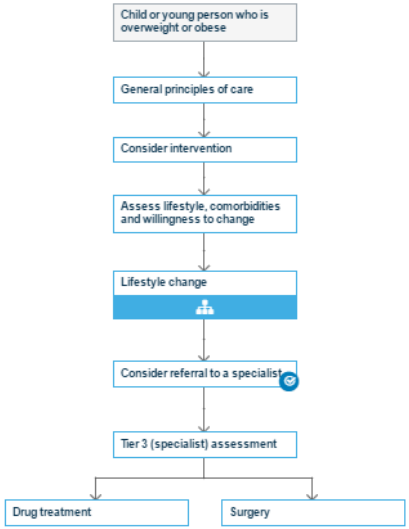
UNIVERSITY OF LEEDS

Typical "As is" pathway



Care pathways are informed by clinical training and working practices within healthcare providers.
Care pathways are implemented, mediated and recorded by health information systems.

Typical "Designed" pathway



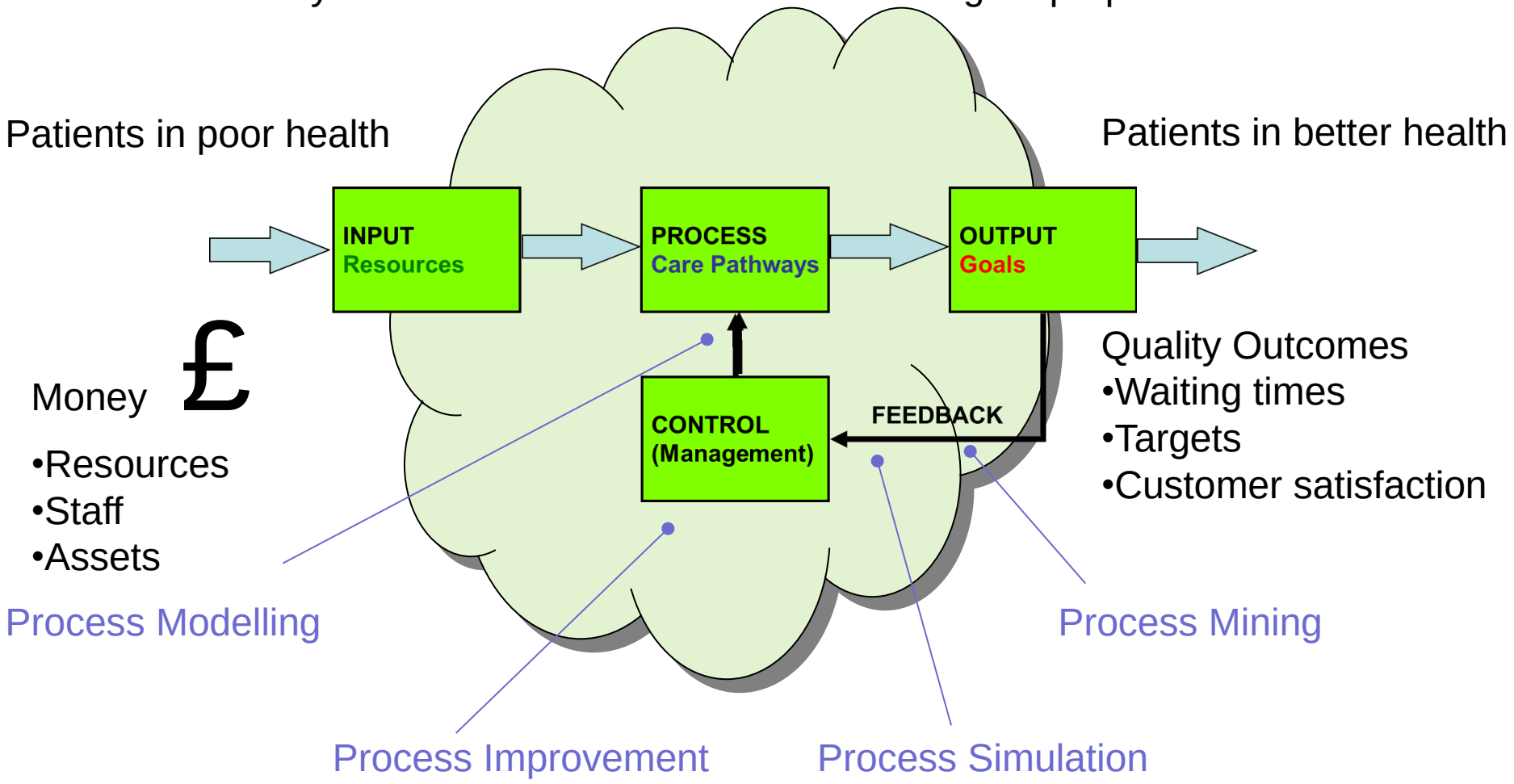
Example: Managing children and young people who are overweight or obese
pathways.nice.org.uk

NB NICE withdrew their comprehensive set of care pathway guidelines in 2022

The challenge
Healthcare is a “Complex System”

UNIVERSITY OF LEEDS

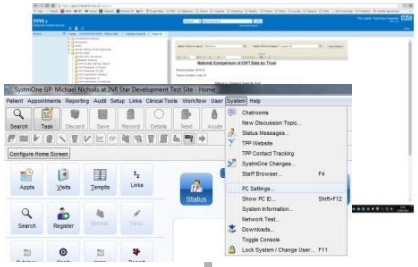
Health Organisations are **complex systems**
They have non-linear behaviour and emergent properties

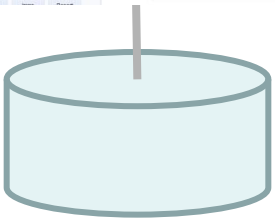


Using EHR data traces

UNIVERSITY OF LEEDS

Electronic Health Record



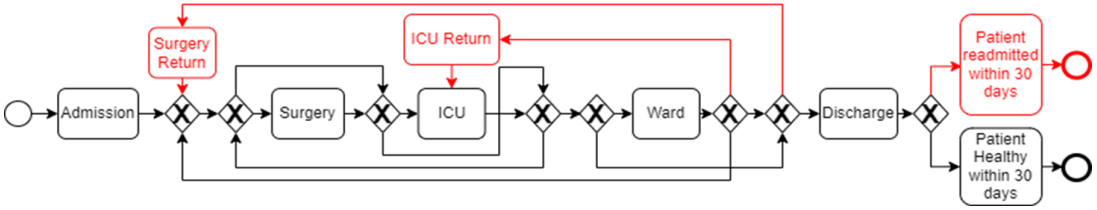


EHR Data

Select data fields of interest (case, activity name, timestamp)
Select activities of interest

Process Mining Event Logs

Case_ID	Time	Activity
1	20120101 16:53	Community Nursing Visit
1	20120108 16:10	Elective inpatient visit
1	20120109 11:30	Community Nursing Visit
1	20120109 11:50	Referral to ASC
1	20120119 15:43	Elective Inpatient Services
1	20120119 05:22	Emergency Inpatient Services
…	…	…



Process Mining
From Spreadsheet to Process Model

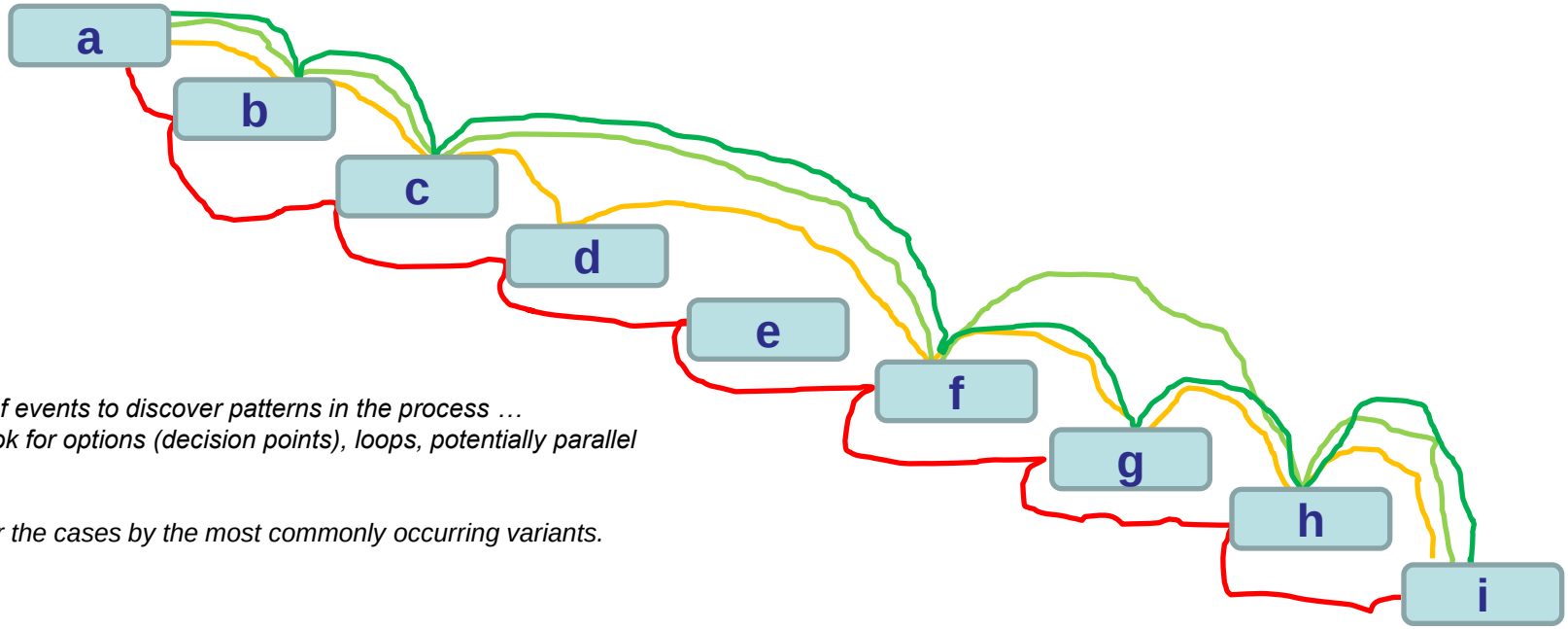
UNIVERSITY OF LEEDS

Process events extracted from a Hospital System Log File

				a		b		c		d		e			f		g		h		i
UniqPat	AE_Si	AE_A	AE_Sex	AE_ArrivalDate		InitialAssessmentTim		AEClinicianSeen		ToSpecialtyTime		FromSpecialtyTi		SpecialtyRefd	BedRequestTime		BedRequestOutcor		AE_LeftDeptTime		PASDischargeDate
1	LGI	77	Male	27/11/2015 14:26		27/11/2015 14:46		27/11/2015 16:43		27/11/2015 18:06		NULL		General Medicine	27/11/2015 15:52		27/11/2015 18:57		27/11/2015 20:29		31/11/2015 18:23:0
2	LGI	22	Male	02/08/2015 16:31		02/08/2015 16:39		02/08/2015 17:26		NULL		NULL		NULL	02/08/2015 17:40		NULL		02/08/2015 19:40		02/08/2015 22:00
3	SJH	70	Male	20/01/2015 19:58		20/01/2015 20:09		20/01/2015 22:15		NULL		NULL		NULL	20/01/2015 21:54		20/01/2015 21:57		21/01/2015 03:22		25/01/2015 12:28
4	SJH	89	Female	02/08/2015 16:41		02/08/2015 16:49		02/08/2015 18:27		02/08/2015 20:34		02/08/2015 20:37		Geriatric Medicine	02/08/2015 19:30		02/08/2015 20:38		02/08/2015 20:41		17/08/2015 16:38
5	SJH	88	Female	02/08/2015 17:24		02/08/2015 17:26		02/08/2015 19:47		NULL		NULL		NULL	02/08/2015 19:01		02/08/2015 21:49		02/08/2015 22:00		16/08/2015 13:03
6	SJH	75	Female	01/11/2015 14:50		01/11/2015 14:57		01/11/2015 15:21		NULL		NULL		NULL	01/11/2015 16:06		01/11/2015 16:46		01/11/2015 17:42		05/11/2015 17:27
7	LGI	70	Female	02/08/2015 19:08		02/08/2015 19:15		02/08/2015 20:08		02/08/2015 20:30		NULL		General Medicine	02/08/2015 21:27		NULL		02/08/2015 22:08		03/08/2015 20:00
8	LGI	74	Female	02/08/2015 19:43		02/08/2015 19:48		02/08/2015 20:27		NULL		NULL		NULL	02/08/2015 20:53		02/08/2015 22:10		02/08/2015 23:24		NULL
9	LGI	98	Female	02/08/2015 21:46		02/08/2015 21:47		02/08/2015 22:03		02/08/2015 22:40		NULL		Geriatric Medicine	02/08/2015 22:57		02/08/2015 23:56		03/08/2015 00:34		07/08/2015 16:40
10	SJH	80	Female	21/11/2015 13:39		21/11/2015 13:44		21/11/2015 13:59		NULL		NULL		NULL	NULL		NULL		21/11/2015 17:06		29/11/2015 00:00

System log files record a sequence of events for different cases

Case01 a,b,c,d,f,g,h,i
Case02 a,b,c,f,h,i
Case03 a,b,c,f,g,h,i
Case04 a,b,c,d,e,f,g,h,i
Case05 a,b,c,f,g,h,i
Case06 a,b,c,f,g,h,i
Case07 a,b,c,d,f,h,i
Case08 a,b,c,f,g,h
Case09 a,b,c,d,f,g,h,i
Case10 a,b,c,h,i



Trace the sequence of events to discover patterns in the process …
… look for options (decision points), loops, potentially parallel

and

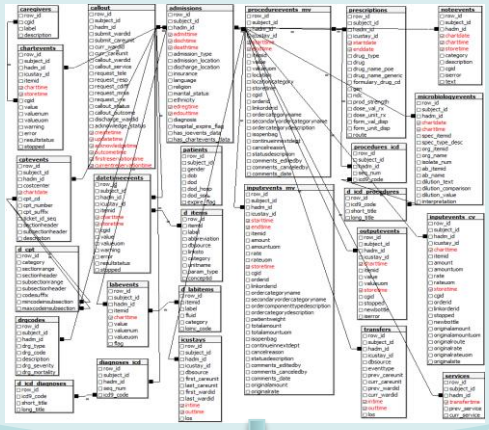
Order the cases by the most commonly occurring variants.

Process Mining

Simple example using the MIMIC-III Open-access dataset

UNIVERSITY OF LEEDS

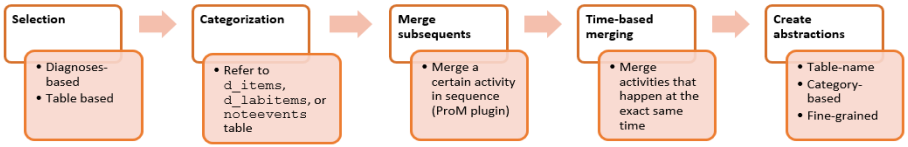
1. The MIMIC-III database (PostgreSQL)



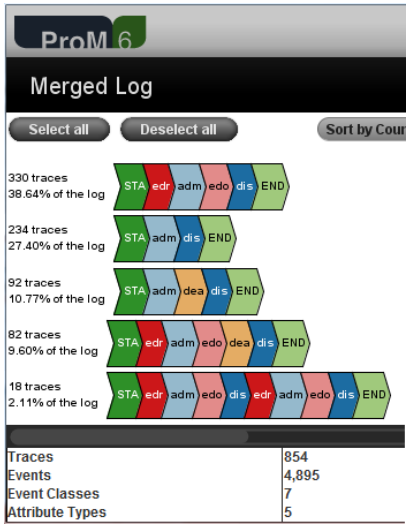
2. Event log creation

	subject_id	hadm_id	activity	charttime
1	2	163353	admit	2138-07-17 19:04:00
2	2	163353	discharge	2138-07-21 15:48:00
3	3	145834	edreg	2101-10-20 17:09:00
4	3	145834	admit	2101-10-20 19:08:00
5	3	145834	edout	2101-10-20 19:24:00
6	3	145834	discharge	2101-10-31 13:58:00
7	4	185777	edreg	2191-03-15 13:10:00
8	4	185777	admit	2191-03-16 00:28:00
9	4	185777	edout	2191-03-16 01:10:00
10	4	185777	discharge	2191-03-23 18:41:00
11	5	178980	admit	2103-02-02 04:31:00
12	5	178980	discharge	2103-02-04 12:15:00
13	6	107064	admit	2175-05-30 07:15:00
14	6	107064	discharge	2175-06-15 16:00:00
15	7	118037	admit	2121-05-23 15:05:00
16	7	118037	discharge	2121-05-27 11:57:00
17	8	159514	admit	2117-11-20 10:22:00
18	8	159514	discharge	2117-11-24 14:20:00
19	9	150750	edreg	2149-11-09 11:13:00
20	9	150750	admit	2149-11-09 13:06:00

3. Preprocessing

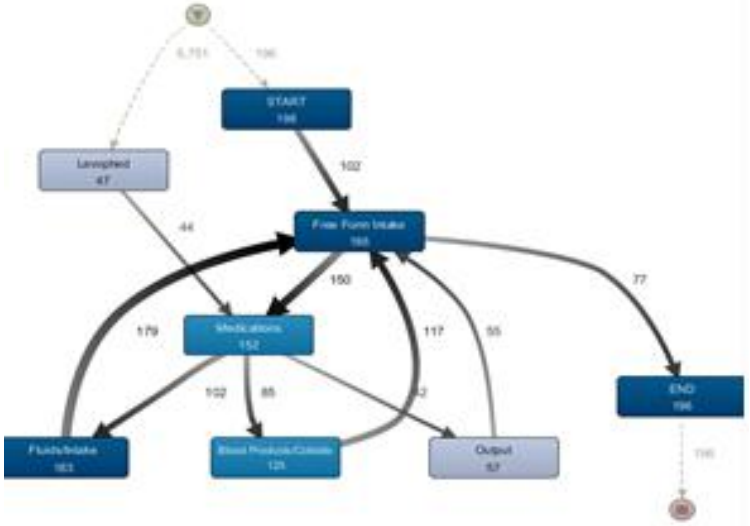


4. Process discovery



admission process variants (using ProM)

5. Process Model



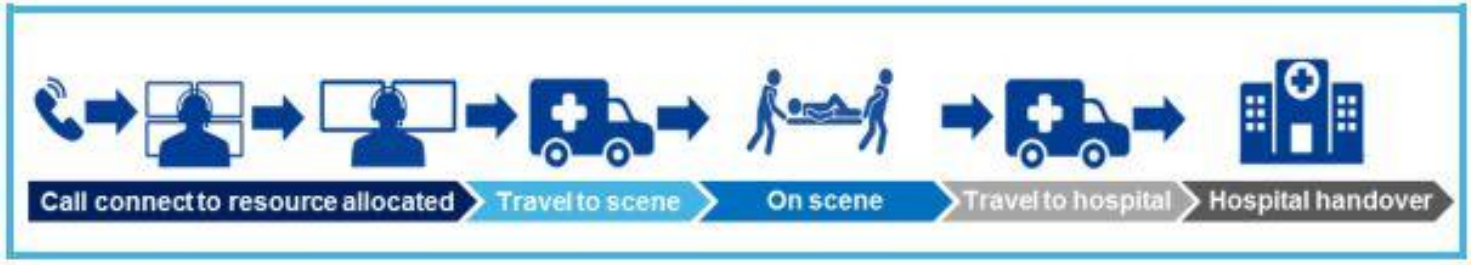
Process model (using Disco)

Kurniati, A.P., Johnson, O., Hogg, D. and Hall, G., 2016, October. *Process mining in oncology: A literature review*. ICICM 2017 (pp. 291-297). IEEE

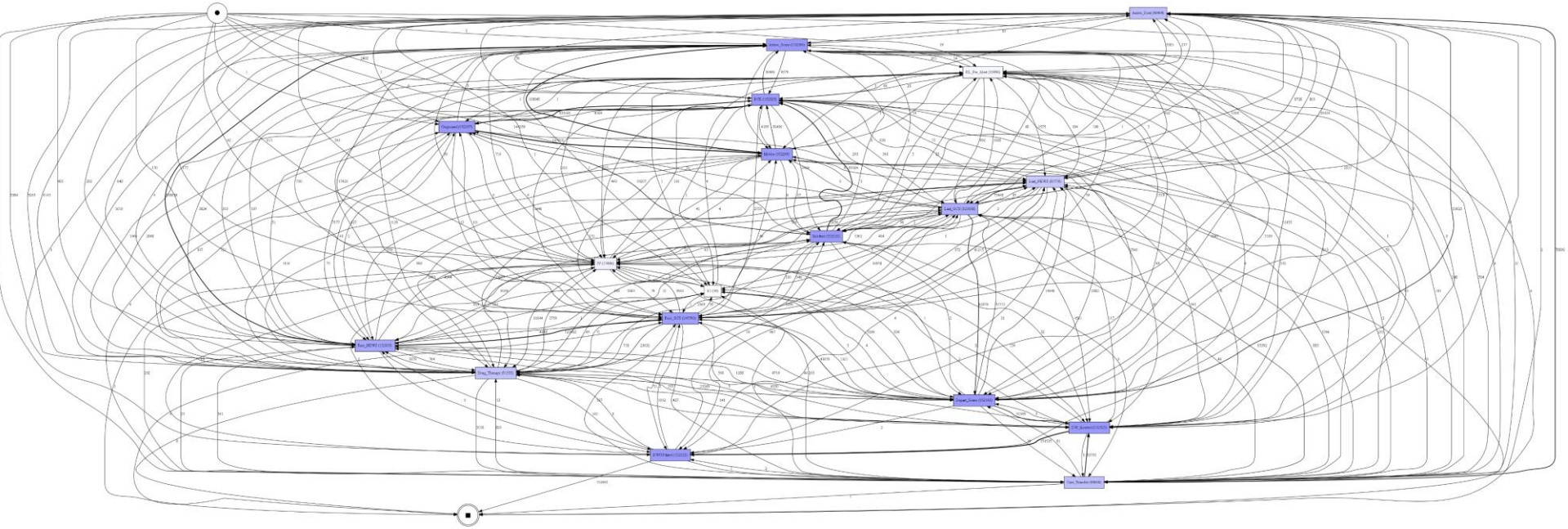
Johnson, A.E., Pollard, T.J., et al., 2016. *MIMIC-III, a freely accessible critical care database*. Scientific data, 3. doi:10.13026/C2XW26

PM Demonstrator Project with NHS England
East Midlands Ambulance Service

UNIVERSITY OF LEEDS



Job Cyle from Public Services Committee (2023), 'Emergency healthcare: a national emergency.', https://committees.parliament.uk/publications/33569/documents/187215/default/[Accessed 25 September 2023



Every Trace Variant for Category 2 Emergency calls (n=152,525)
Alex Coles (Leeds AI Medical CDT) with Owen Johnson. Project for NHS England 2024.

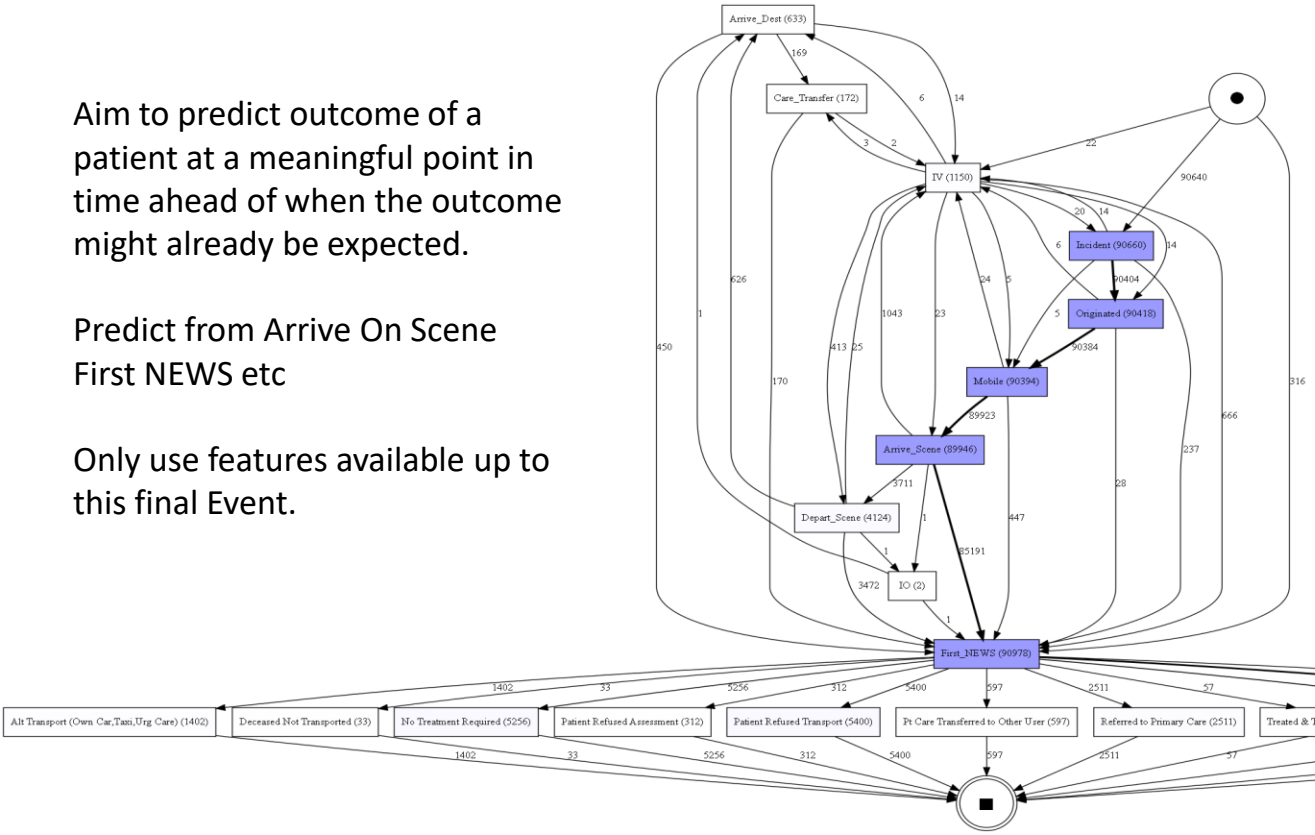
Applying Machine Learning to Predict Care Pathway Outcome

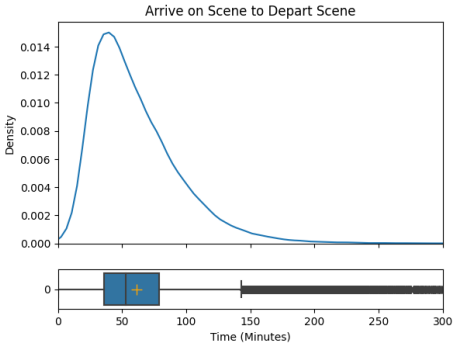
UNIVERSITY OF LEEDS

Aim to predict outcome of a patient at a meaningful point in time ahead of when the outcome might already be expected.

Predict from Arrive On Scene First NEWS etc

Only use features available up to this final Event.





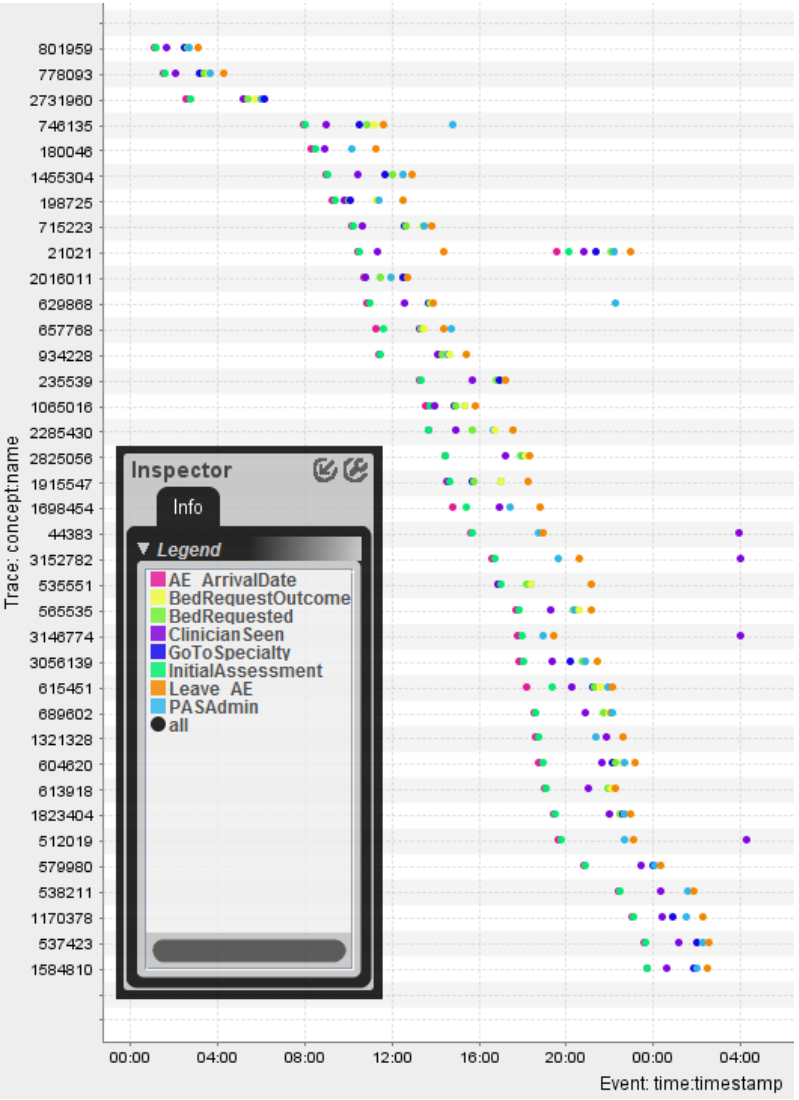
Decision Tree Feature Importances

Feature	Importance
Initial Clinical Category of Call	0.29
First Recorded National Early Warning Score (NEWS)	0.23
Age of Patient	0.17
Hour of Arrival on Scene	0.10
Index of Multiple Deprivation (IMD) Decile at Incident Location	0.075
Day of the Week	0.065
Highest Qualification on Scene	0.038

Alex Coles (Leeds AI Medical CDT) with Owen Johnson. Project for NHS England 2024.

The case for data intimacy
Leeds Hospital A&E Department

UNIVERSITY OF LEEDS

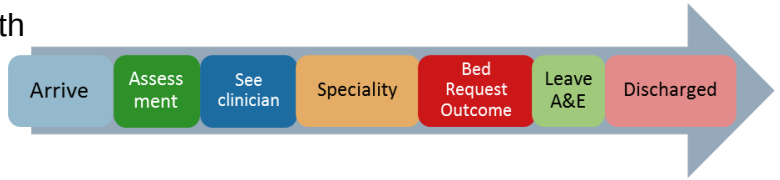


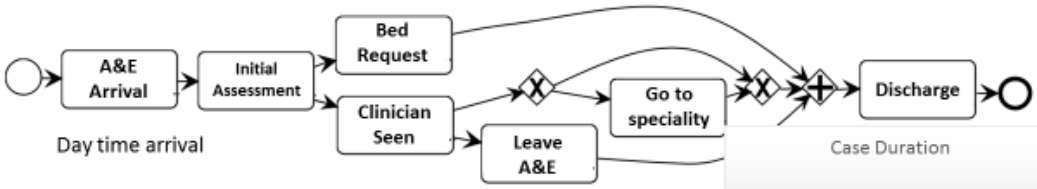
A&E Time

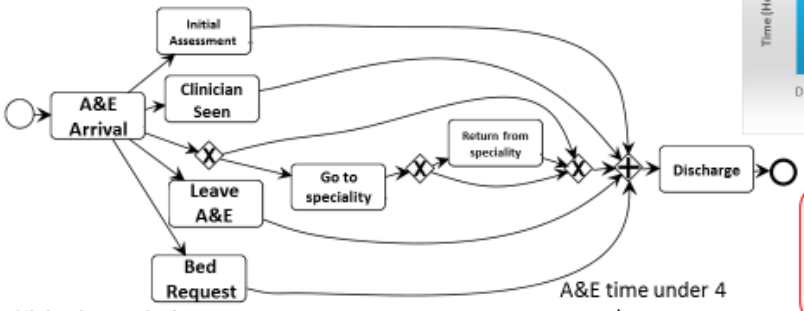
67% of episodes are under 4 hours in length

Min: Instant | Median: 3.9 hours | Mean: 4.3 hours | Max: 17 hours

Most common path
(49.6% of cases)







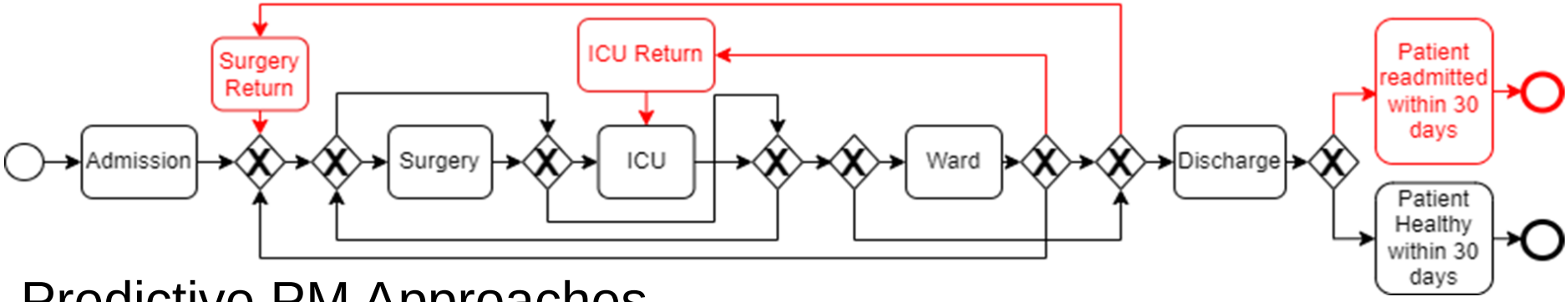
Is there a different process during the night?

De-identified data from A&E data for patients admitted to an acute medicine ward, 14,000 admissions - July 2014 – July 2015

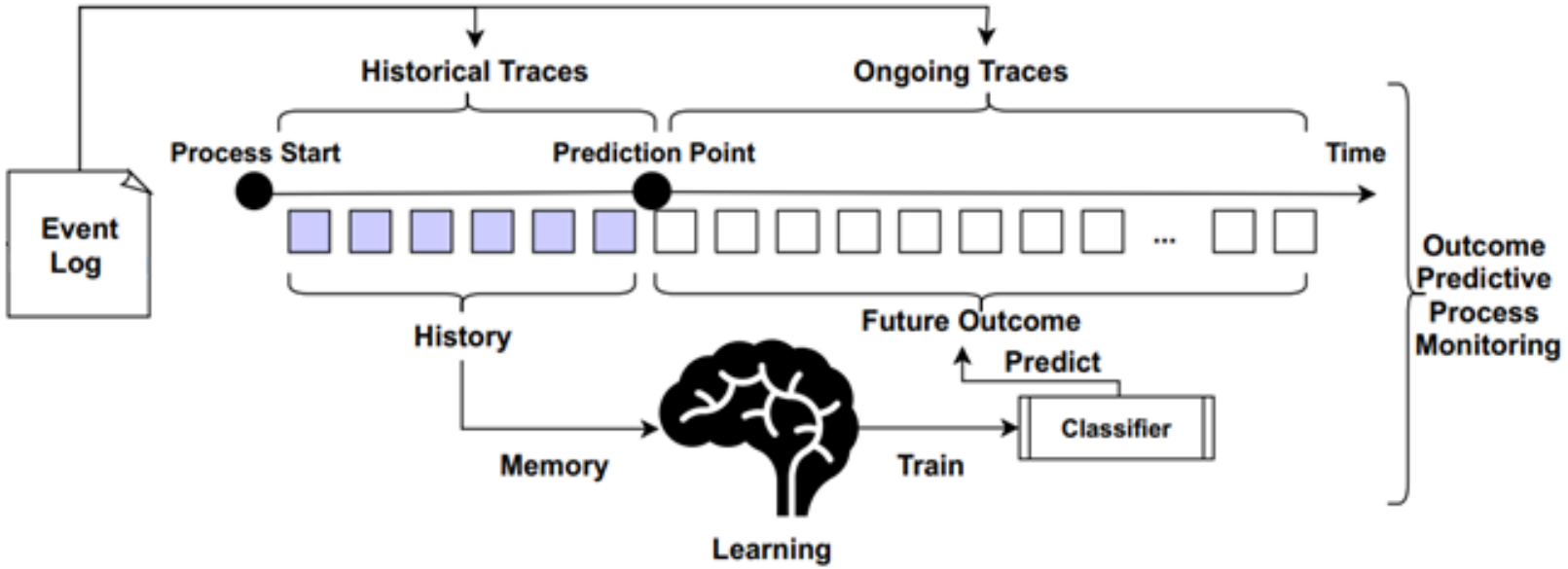
Care Pathways in the NHS
Mining and Prediction of “Left Shifts”

UNIVERSITY OF LEEDS

Left Shifts



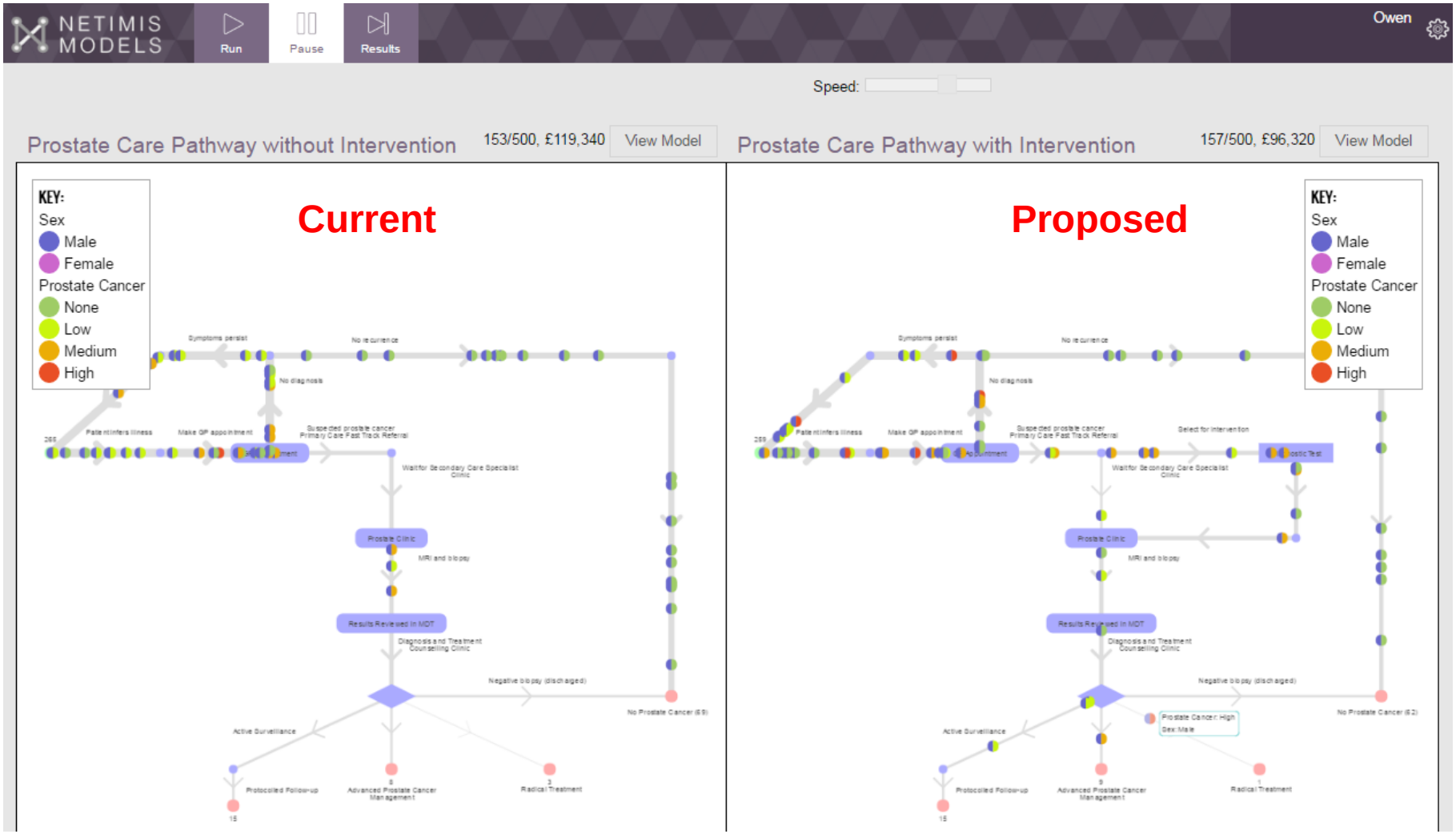
Predictive PM Approaches



Chen, Qifan, et al. "Outcome-Oriented Predictive Process Monitoring to Predict Unplanned ICU Readmission in MIMIC-IV Database." (2022).

Process Improvement
Cost/ Benefit Evaluation of Options

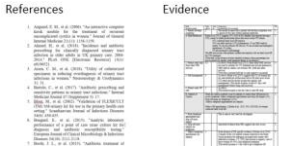
UNIVERSITY OF LEEDS



Industry project for USA Genetics company - Process simulation of a prostate cancer intervention. Cathy Tomlinson and Owen Johnson (Bekker, £166k, 2017-2019)

Process Mining and Process Simulation

UNIVERSITY OF LEEDS

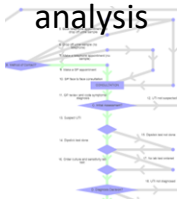


Real-life care

Observed process

Process Simulation

Process simulations and "what if" analysis



IT Systems

Syntheti c event log

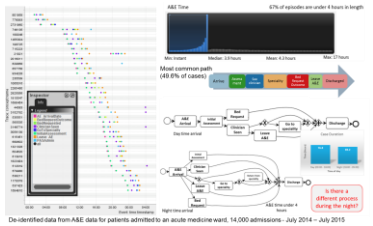
Process model

Detailed Process analytics and metrics

Data

Real even t log

Process Mining



ClearPath
Process Mining in Healthcare Methodology

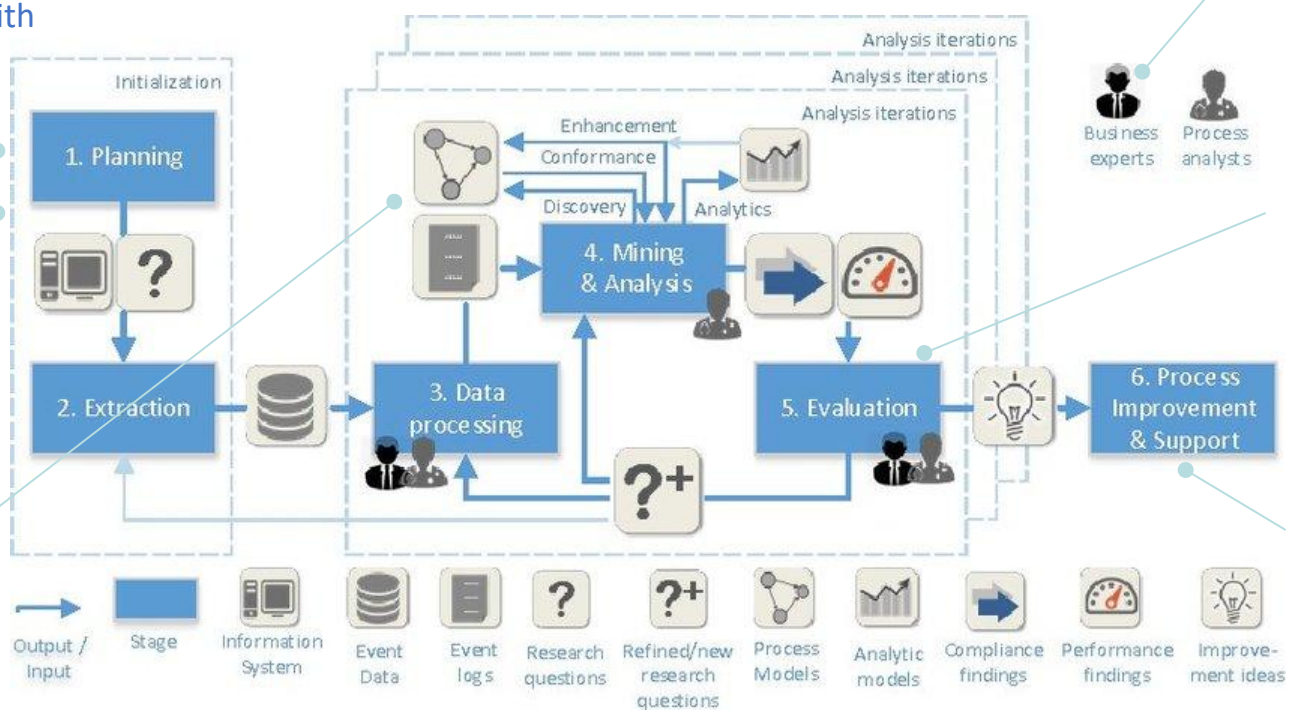
UNIVERSITY OF LEEDS

ClearPath adapts the industry-standard PM2 method to the healthcare domain and care pathways

ClearPath starts with well-recognised patterns

ClearPath incorporates literature review data

ClearPath also uses “plain old business process re-engineering” (PO-BPR) methods



ClearPath adds a Clinical Reference group

ClearPath can include modelling local variants, population models, activity costs and times.

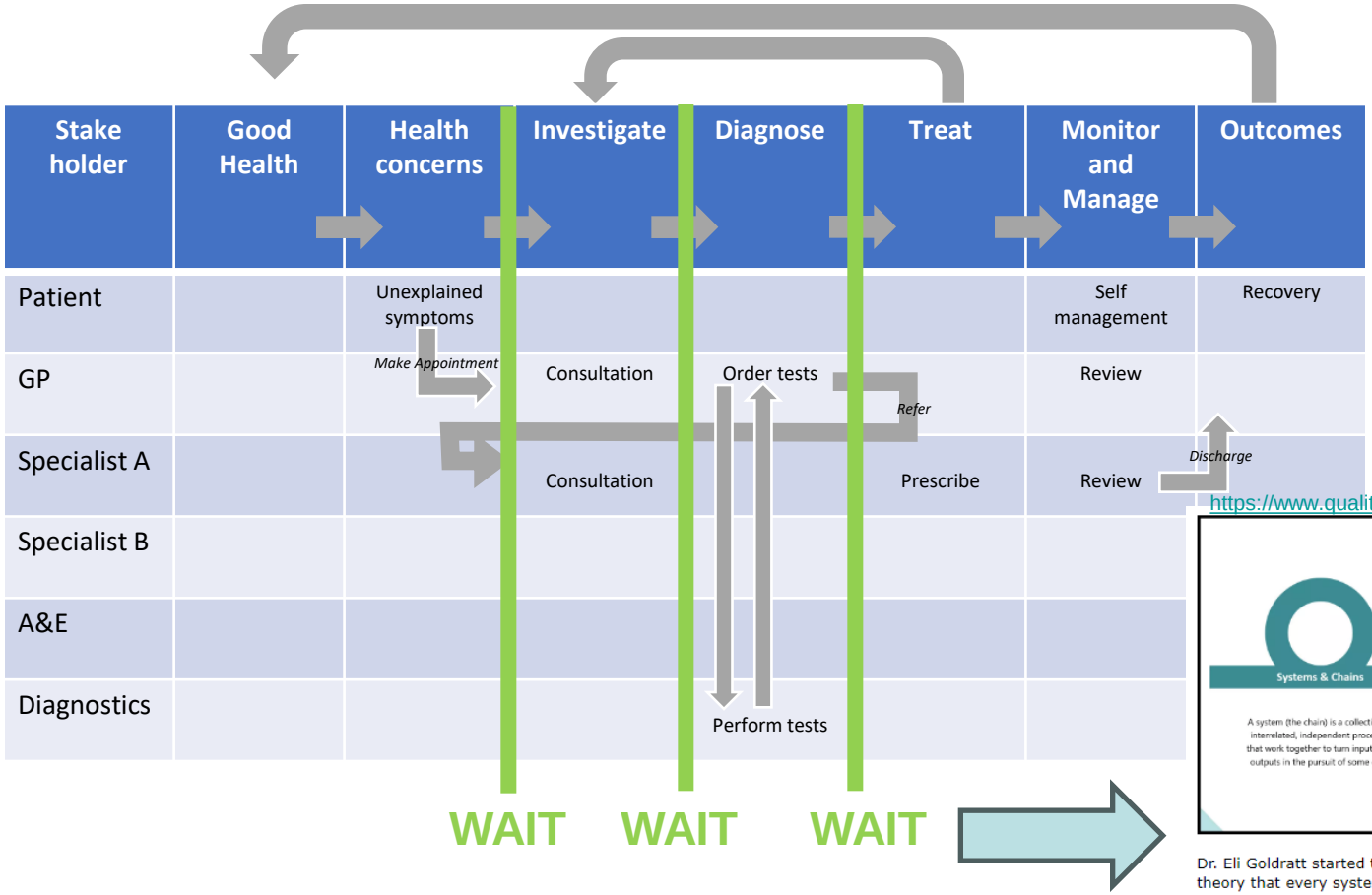
ClearPath extends to process simulation for “what if” scenarios

Johnson, O.A., Ba Dhafari, T., Kurniati, A., Fox, F. and Rojas, E., 2019. The clearpath method for care pathway process mining and simulation. In *Business Process Management Workshops: BPM 2018 International Workshops, Sydney, NSW, Australia, September 9-14, 2018, Revised Papers 16* (pp. 239-250). Springer International Publishing.

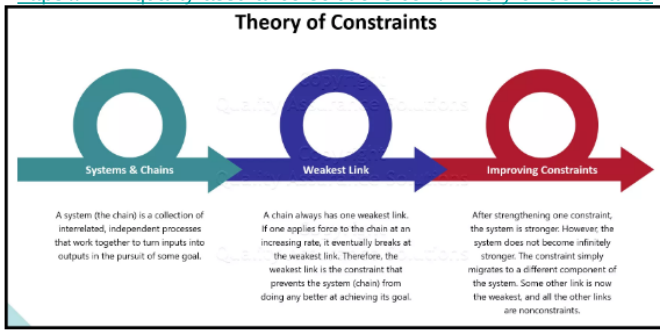
Care Pathways in the NHS follow typical patterns

UNIVERSITY OF LEEDS

Most care pathways will follow one or more standard template patterns
Linear, Cycle, Referral, Sub-process, Parallel process.
Waiting occurs when the process is constrained by resources.



https://www.quality-assurance-solutions.com/Theory-of-Constraints

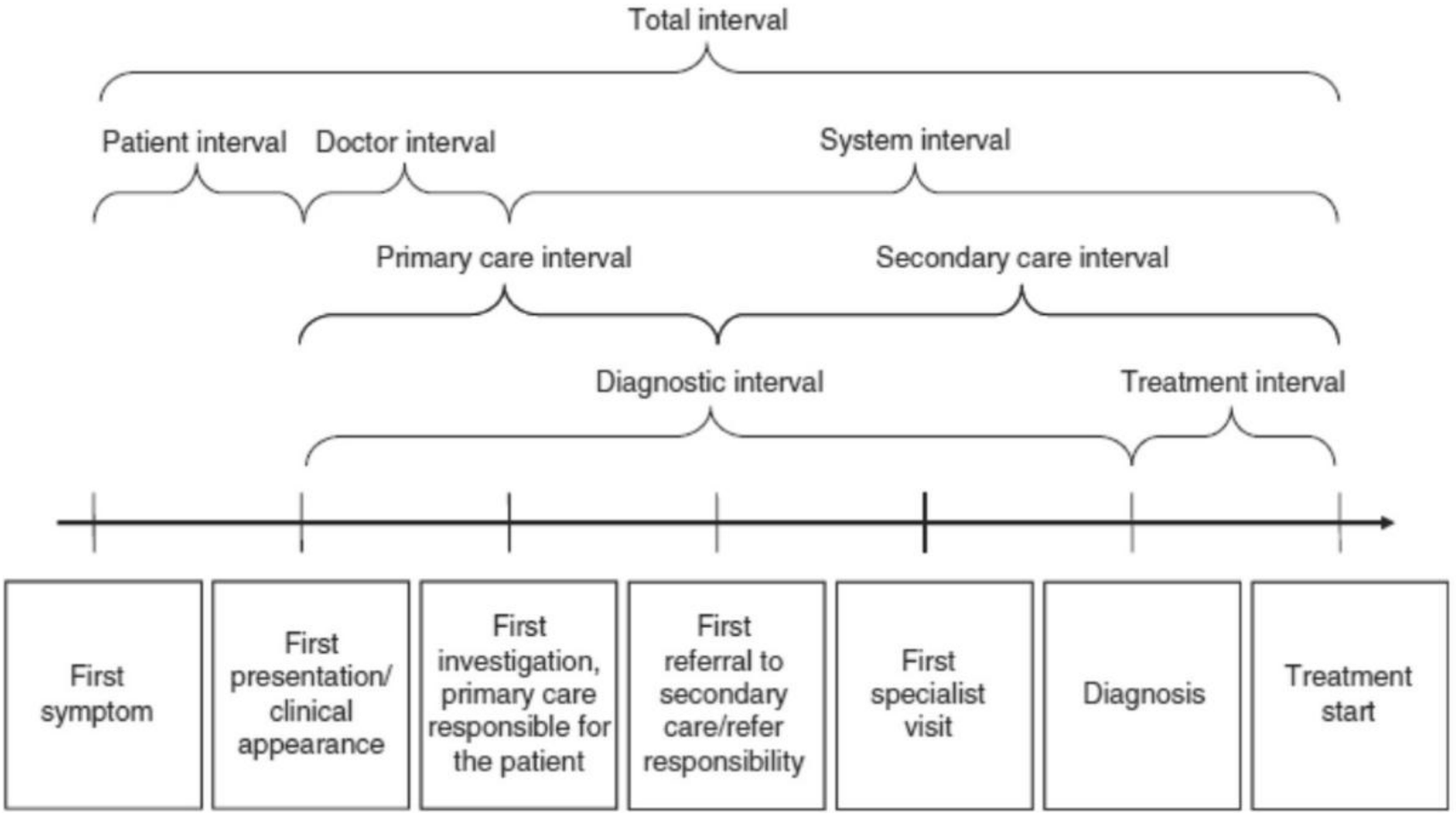


Dr. Eli Goldratt started the theory of constraints (TOC). He based this management theory that every system has at least one constraint limiting it from getting more of what it strives for. If this were not true, then the system would produce infinite output.

Patterns in Care Pathways

A framework for medical process activities and intervals

UNIVERSITY OF LEEDS



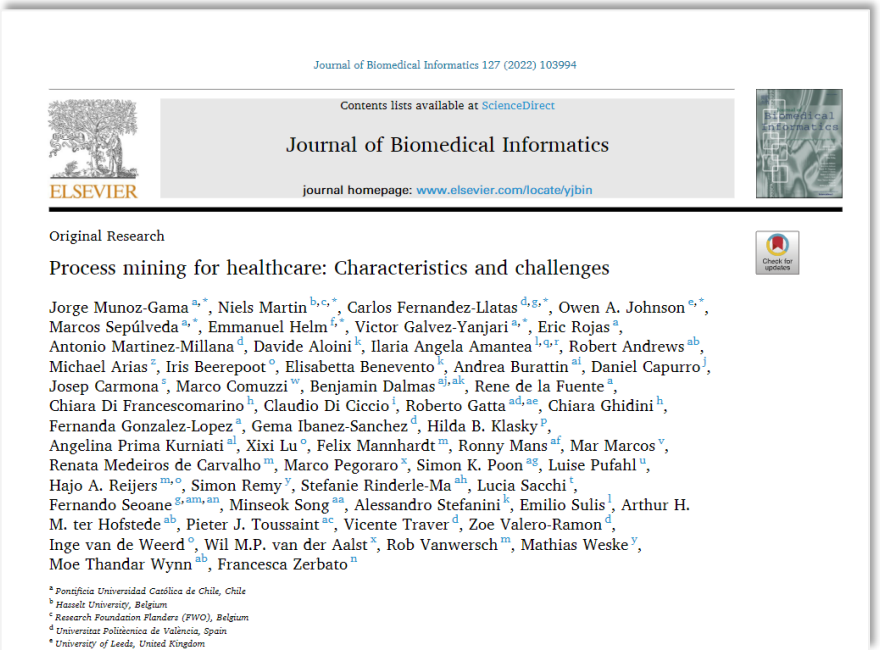
Weller, D., Vedsted, P., Rubin, G., Walter, F.M., Emery, J., Scott, S., Campbell, C., Andersen, R.S., Hamilton, W., Olesen, F. and Rose, P., 2012. The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. British journal of cancer, 106(7), pp.1262-1267.

https://pubmed.ncbi.nlm.nih.gov/22415239

The Process Mining for Healthcare Manifesto

UNIVERSITY OF LEEDS

Based on a two-day brainstorming event in Hasselt, Belgium (July 2019)
Identified ten Characteristics of healthcare that make Process Mining in health different
And ten Challenges for future research…
.



Journal of Biomedical Informatics 127 (2022) 103994

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

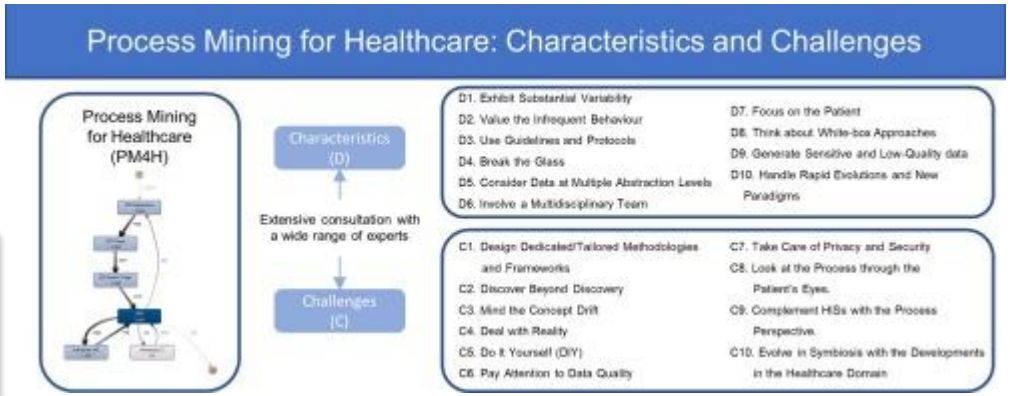
journal homepage: www.elsevier.com/locate/yjbin

Original Research

Process mining for healthcare: Characteristics and challenges

Jorge Munoz-Gama [a,*], Niels Martin [b,c,*], Carlos Fernandez-Llatas [d,g,*], Owen A. Johnson [e,*], Marcos Sepúlveda [a,*], Emmanuel Helm [f,*], Victor Galvez-Yanjari [a,*], Eric Rojas [a], Antonio Martinez-Millana [d], Davide Aloini [k], Ilaria Angela Amantea [l,q,r], Robert Andrews [ab], Michael Arias [z], Iris Beerepoot [o], Elisabetta Benevento [k], Andrea Burattin [ai], Daniel Capurro [j], Josep Carmona [s], Marco Comuzzi [w], Benjamin Dalmas [aj,ak], Rene de la Fuente [a], Chiara Di Francescomarino [h], Claudio Di Ciccio [i], Roberto Gatta [ad,ae], Chiara Ghidini [h], Fernanda Gonzalez-Lopez [a], Gema Ibanez-Sanchez [d], Hilda B. Klasky [p], Angelina Prima Kurniati [al], Xixi Lu [o], Felix Mannhardt [m], Ronny Mans [af], Mar Marcos [v], Renata Medeiros de Carvalho [m], Marco Pegoraro [x], Simon K. Poon [ag], Luise Pufahl [u], Hajo A. Reijers [m,o], Simon Remy [y], Stefanie Rinderle-Ma [ah], Lucia Sacchi [t], Fernando Seoane [g,am,an], Minseok Song [aa], Alessandro Stefanini [k], Emilio Sulis [l], Arthur H. M. ter Hofstede [ab], Pieter J. Toussaint [ac], Vicente Traver [d], Zoe Valero-Ramon [d], Inge van de Weerd [o], Wil M.P. van der Aalst [x], Rob Vanwersch [m], Mathias Weske [y], Moe Thandar Wynn [ab], Francesca Zerbato [n]

[a] Pontificia Universidad Católica de Chile, Chile
[b] Hasselt University, Belgium
[c] Research Foundation Flanders (FWO), Belgium
[d] Universitat Politècnica de València, Spain
[e] University of Leeds, United Kingdom



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

Recommendations

UNIVERSITY OF LEEDS

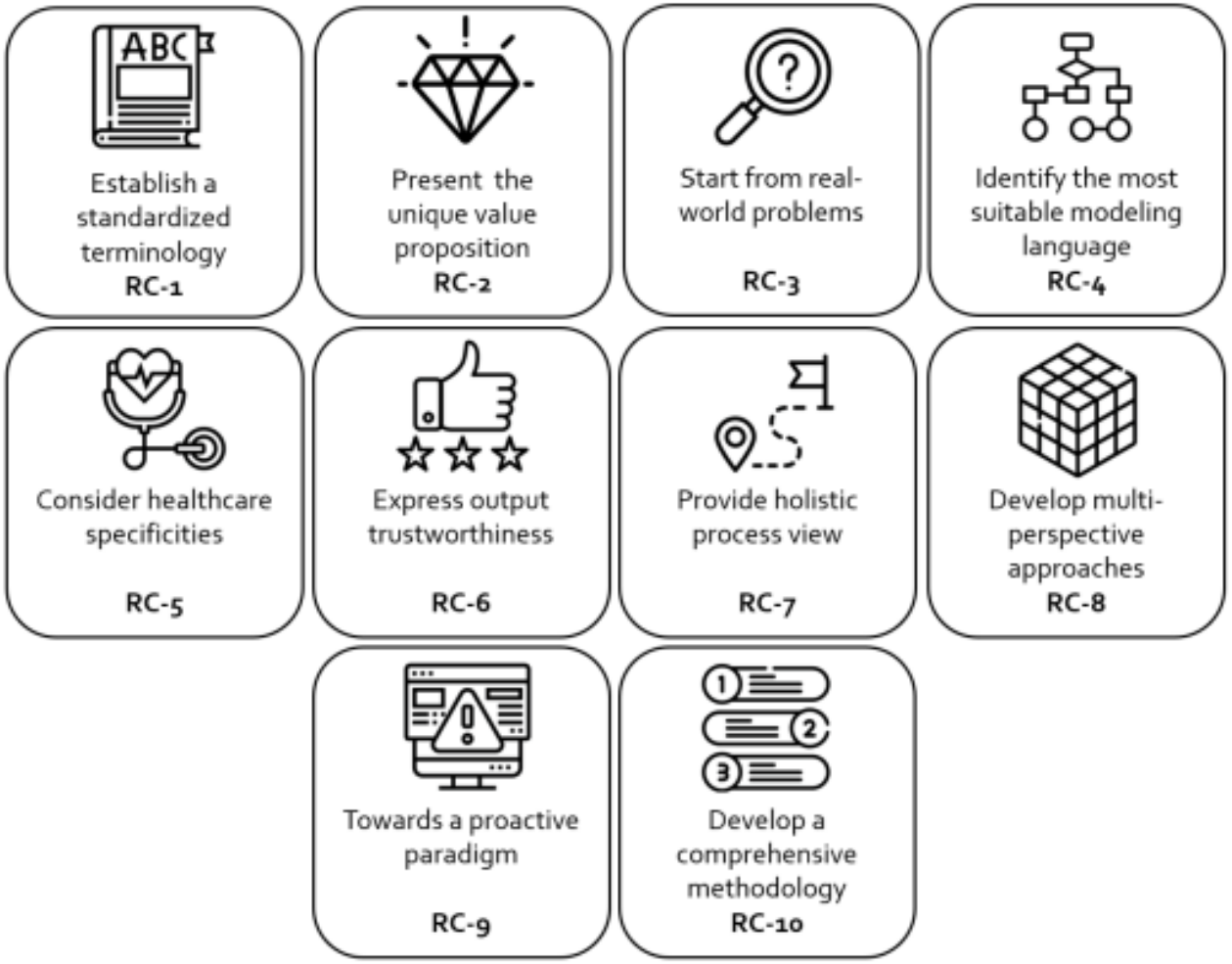


Figure 2: Overview of recommendations for process mining researchers and the research community

Summary:
NO AI without Process Mining, Please.

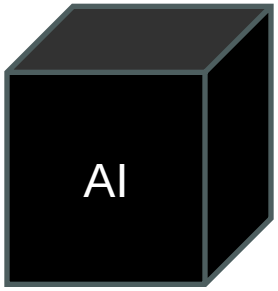
UNIVERSITY OF LEEDS

"Huge public rollout of AI"

Current → **Future**

DEVELOPMENT CHALLENGE
What AI?

What data will it use?
How was it trained?
What risks and bias?



IMPLEMENTATION CHALLENGE
Where (in the process) will it be implemented?

What are the real-world implications?
How will the users be integrated?
What risks and bias?

Images: ChatGPT 4.0

Tools for Process Mining in Healthcare

UNIVERSITY OF LEEDS

Tools for ETL (Extract, Transform, Load)

SQL – to extract data from relational databases

Programming languages e.g. Python, R, Java, C++ etc - to transform into event logs

Commercial Tools

Fluxicon Disco - https://fluxicon.com/disco

Celonis - www.celonis.com

Signavio - https://www.signavio.com/process-mining

More details at https://www.gartner.com/reviews/market/process-mining/vendor/celonis/alternatives

Free Tools

ProM - https://www.promtools.org

BupR - https://www.bupar.net

PM4Py - https://pm4py.fit.fraunhofer.de

More details at https://aimultiple.com/process-mining-software

Process Mining in Healthcare – Opportunities and Challenges

UNIVERSITY OF LEEDS

An introduction to process mining in health and how it can help drive innovation in care, lower costs and improve patient outcomes

- The NHS is under pressure to improve its clinical and operational processes through better digital health and emerging AI.
- Process mining combines data science and process science methods for data-driven process improvement.
- The NHS is rich in data but using it effectively requires state-of-the-art tools, methods and skills.
- This talk will demonstrate how process mining can be used to improve NHS care pathways
- It will present the ClearPath method for process mining in healthcare developed at the University of Leeds
- … and discuss the opportunities and challenges in practice.

School of Computer Science
FACULTY OF ENGINEERING

UNIVERSITY OF LEEDS

Process Mining in Healthcare Opportunities and Challenges

PRSB Professional Record Standards Body

BCS Faculty of Health and Care *and*
Professional Record Standards Body
20th March 2025
17.30-18.30

Owen Johnson

o.a.johnson@leeds.ac.uk

Associate Professor – Digital Health

Director of Impact, School of Computer Science

Co-Director, Leeds CDT for AI in Medical Diagnosis and Care

Programme Manager, MRes in Data Science for Healthcare (NHS England)