MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190261

Process Mining in Primary Care: Avoiding Adverse Events Due to Hazardous Prescribing

Richard Williams[a], Darren M. Ashcroft[a], Benjamin Brown[a], Eric Rojas[b], Niels Peek[a], Owen Johnson[c]

[a] NIHR Greater Manchester Patient Safety Translational Research Centre, University of Manchester, Manchester, UK
[b] Internal Medicine Department. School of Medicine. Pontificia Universidad Católica de Chile
[c] NIHR Yorkshire and the Humber Patient Safety Translational Research Centre, University of Leeds, Leeds, UK

Abstract

Process mining helps healthcare professionals understand processes within healthcare. While often used in secondary care, there is little work in process mining using primary care data. Serious adverse events that result from hazardous prescribing are common and costly. For example, non-steroidal anti-inflammatory drugs (NSAIDs) and antiplatelets can cause gastro-intestinal bleeds (GiBs). Prescribing typically occurs during primary care; therefore we used this setting to attempt process mining. We extracted events (drug started, drug stopped, GiB) for understanding three prescribing pathways, and applied process mining. We found NSAIDs are often short-term prescriptions whereas antiplatelets are often long-term. This perhaps explains our finding that co-prescription of gastro-protection is more prevalent for antiplatelets than NSAIDs. We identified reasons why primary care data is harder to process mine and proposed solutions. Process mining primary care data is possible and likely useful for improving patient safety and reducing costs.

Keywords:
Data mining, Patient safety, Primary care.

Introduction

Process mining describes a collection of methods for extracting information about processes from event logs [1]. There are three distinct stages: detecting the underlying process from the event logs (process discovery); identifying deviations from what was expected (conformance checking); and generating suggestions for redesigning and improving the processes (enhancement) [1]. Process mining adds a temporal dimension to standard data mining methods. Originally applied to business processes, more recently has been applied to other domains including healthcare. In a recent literature review, we showed that while process mining within secondary and tertiary care has become more common, there is almost no work within primary care [2].

Patient safety is fundamental to healthcare systems. Within UK primary care this is true for medication prescribing where life-threatening errors appear in 1 in 550 prescriptions [3]. A recent economic analysis showed that: adverse drug reactions (ADRs) cost the NHS up to £1.6 billion a year; more than one third of ADR related hospital admissions are caused by non-steroidal anti-inflammatory drugs (NSAIDs), antiplatelets and anticoagulants; and half of the deaths associated with primary care ADRs are due to gastro-intestinal bleeds (GiBs) [4]. Studying the relationship between the prescribing practice of NSAIDs and antiplatelets with ADRs including GiBs is therefore important.

In the UK there are several large databases of coded primary care records available for research [5]. While the quality of coding may not be universally high [6], all practice-based prescribing in primary care is electronic so therefore this would be a suitable place to attempt to apply process mining.

While the epidemiology of hazardous prescribing in primary care has been extensively studied using large electronic health record databases, to date little is known about the typical processes that lead to such prescriptions. To design effective interventions for reducing hazardous prescribing, it is essential to get a better understanding of these processes. This could lead to better decision support systems for prescribers, and ultimately improve patient safety and reduce cost by reducing the number of ADRs.

Our objective was therefore to process mine UK primary care data to explore the relationship between the prescribing of NSAIDs, antiplatelets and the adverse outcome of gastro-intestinal bleeds.

Methods

A process model is a graphical representation of a process showing the events and how they interrelate via directed edges. Process discovery is the extraction of a process model from an event log via the application of an algorithm. There are many algorithms with various strengths and weaknesses. For example the α-algorithm is simple and therefore easy to undersand, but it does not deal well with noisy event logs which are typical of real world processes [1]. Heuristic Miner and, in particular, Fuzzy Miner are better able to deal with this noise [7]. Here we focus on process discovery to prove that process mining can be applied to primary care data, and use the Fuzzy Miner to best handle the messiness of routinely collected health data.

Anonymised patient data was obtained from the Salford Integrated Record (SIR); a data warehouse with contributions from 43 general practices in Salford, UK (population 0.25M). All coded data, including diagnoses and medications, for patients who have not opted out (1.5% opt outs) was available to extract from a SQL Server database. The earliest records are historic diagnoses from the 1940s, but the bulk of the data collection is from 2000 onwards. Approval was granted by the SIR governance board and all data was obtained pseudonymised (random identifier, no name, year of birth instead of age, geographic region instead of address).

A review by Spencer *et al.* [8] identified 56 prescribing safety indicators for use in primary care to improve patient safety. They each try to prevent a particular adverse outcome through safer prescribing, e.g., patients with chronic kidney disease should not be prescribed an NSAID because of the increased risk of acute renal failure. A subset of these indicators are

included in: electronic audit and feedback initiatives such as the national PINCER [9] rollout and the SMASH intervention [10]; and clinical decision support systems such as OptimizeRx [11]. We selected three prescribing safety indicators for further analysis that focus on NSAIDs and anticoagulants, and are designed for preventing GiBs in cohorts of patients at increased risk such as the elderly and those with a history of peptic ulceration. The indicators and the descriptions used in our analyses are provided in Table 1.

Table 1– Prescribing Safety Indicators for Preventing GiBs

Id	Short name	Description
I1	Age≥65 + NSAID	Patients aged 65 and over who are prescribed an NSAID should also be prescribed a gastro-protective medication (GPM).
I2	Pep + NSAID	Patients with a history of peptic ulceration who are prescribed an NSAID should also be prescribed a GPM.
I3	Pep + Antiplatelet	Patients with a history of peptic ulceration who are prescribed an antiplatelet should also be prescribed a GPM.

Prescription events are recorded automatically in a patient's record; however, the stopping of medication is not recorded. We have previously developed an algorithm to convert these prescription events into more meaningful events such as when a drug is started and stopped, and when a dose is changed [12]. This process is done by evaluating: the date of the prescription; the amount prescribed; and the rate at which it is consumed. For the medications of interest (NSAIDs, GPMs, and antiplatelets), we extracted the start and stop events.

For each indicator, we developed queries that would extract the patient data. First, we defined the cohort of patients from an initial event. For indicator I1, it was when a patient turned 65, and for I2 and I3, it was the first instance of a peptic ulceration. Next, we extracted the start and stop events for all medications of interest. Finally, we extracted other relevant events: GiB and peptic ulceration classified as either the first bleed, or a subsequent bleed; patient turned 65; and patient died. Clinical code sets were constructed for each event of interest [13].

The initial output of our medication algorithm gives the start and stop events for individual active ingredients so, for example, for two different NSAIDs we would have two different start events. However, we are only interested in whether any NSAID (or other medication) is started if the patient is not already taking an existing one. Similarly, when a drug is stopped it is only relevant if the patients are then not taking any other drug of the same type. An additional processing script was therefore required to produce the final event log. This additional data processing was done using JavaScript and nodejs [14].

The data was extracted on 2nd November 2018 and process mining was performed by the lead author. Process mining was performed using Fluxicon Disco (academic licence) [15] on a Dell XPS 15 laptop running Windows 10. All clinical code sets and processing code is at https://zenodo.org/record/1493640.

Results

The demographic information for the patient cohorts for each indicator are displayed in Table 2. The median duration time, interquartile range, and number of transitions between events are shown in Tables 3-5. For example, in Table 3, the event "Bleed" immediately followed the event "Age 65" in the event logs 390 times, with a median transition time of 49 (IQR [19,112]) months. The process mining diagrams extracted from Disco are shown in Figures 1-3. The numbers on the nodes in the diagrams represent the number of times each event occurred, while the edge numbers are the number of times the target event directly followed the source event.

Table 2– Patient Characteristics for Each Cohort and the Population of Salford. Values are n (%) unless otherwise specified.

Demographic	I1		I2, I3		All patients	
# of patients	38,936		3,477		270,412	
Age						
Mean (SD)	76	(8)	66	(16)	37	(23)
Sex						
Female	20,633	(53%)	1,238	(36%)	131,935	(49%)
Male	18,303	(47%)	2,239	(64%)	138,473	(51%)
Ethnicity						
White	16,291	(42%)	1,444	(42%)	96,696	(36%)
Other	643	(2%)	139	(4%)	24,124	(9%)
Unknown	22,002	(57%)	1,894	(54%)	149,592	(55%)
Deprivation [16] quintiles						
1st (highest)	15,023	(39%)	1,618	(47%)	115,593	(43%)
2nd	8,126	(21%)	694	(20%)	56,284	(21%)
3rd	7,536	(19%)	569	(16%)	44,770	(17%)
4th	3,533	(9%)	258	(7%)	19,591	(7%)
5th (lowest)	4,320	(11%)	269	(8%)	19,561	(7%)

Table 3– Median Duration in Months of Transitions between Key Events for Indicator I1. IQR in [square] brackets. Number of transitions in (round) brackets. NSAID – non-steroidal anti-inflammatory drug, GPM – gastro-protective medication.

Next event / Event	Bleed	NSAID (no GPM)	NSAID (GPM)	NSAID Stopped	GPM Started	GPM Stopped
Age 65	49 [19,112] (390)	33 [11,74] (10925)	17 [6,39] (1385)	9 [2,28] (1684)	48 [15,111] (9366)	12 [3,37] (1617)
Bleed	1 [0,5] (523)	14 [3,47] (48)	22 [7,46] (75)	1 [0,3] (54)	0 [0,3] (582)	10 [2,34] (276)
NSAID (no GPM)	6 [1,36] (34)			1 [1,3] (24106)	0 [0,0] (9478)	
NSAID (GPM)	2 [1,32] (11)			1 [1,3] (8733)		1 [1,5] (566)
NSAID Stopped	20 [5,57] (292)	11 [5,26] (14923)	10 [5,22] (4553)		18 [5,46] (6913)	0 [0,5] (7769)
GPM Started	5 [1,23] (483)		9 [3,27] (3966)	1 [1,2] (6548)		1 [1,5] (28868)
GPM Stopped	13 [3,38] (231)	11 [5,25] (5719)		1 [1,4] (2724)	6 [2,14] (21444)	

For indicator I1 there were 45,479 NSAID start events. Of these, 9,981 (22%) were for patients already prescribed a GPM. A further 9,478 (21%) then started GPM at a median duration of 0 days suggesting co-prescription. However, 24,106 (53%) NSAID start events were followed by an NSAID stop event at a median duration of 1 month (IQR [1, 3] months), suggesting a short term prescription without co-prescription of a GPM.

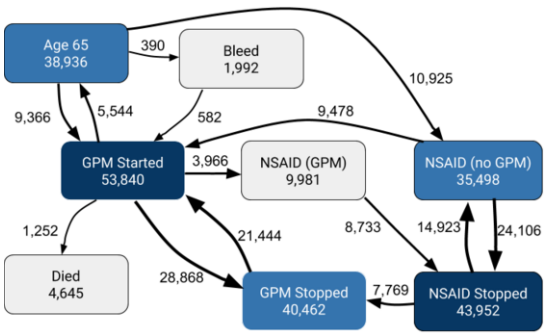

Figure 1– Process Diagram for Indicator I1. NSAID – non-steroidal anti-inflammatory drug, GPM – gastro-protective medication.

Table 4– Median Duration in Months of Transitions between Key Events for Indicator I2. IQR in [square] brackets. Number of transitions in (round) brackets. NSAID – non-steroidal anti-inflammatory drug, GPM – gastro-protective medication.

Next event / Event	Bleed	NSAID (no GPM)	NSAID (GPM)	NSAID Stopped	GPM Started	GPM Stopped
Initial Bleed	3 [0,60] (603)	154 [46,283] (564)			2 [1,71] (2102)	
Bleed	2 [1,7] (752)	57 [17,107] (125)	18 [4,52] (88)	1 [0,8] (21)	1 [0,24] (729)	5 [1,24] (299)
NSAID (no GPM)	1 [0,6] (14)			1 [1,2] (2388)	0 [0,1] (861)	
NSAID (GPM)	5 [1,27] (11)			1 [1,3] (1416)		2 [1,5] (112)
NSAID Stopped	15 [4,38] (157)	14 [7,30] (1437)	11 [6,25] (695)		13 [4,36] (910)	1 [0,11] (912)
GPM Started	4 [1,19] (501)		13 [4,33] (819)	1 [1,3] (630)		2 [1,7] (6884)
GPM Stopped	9 [3,27] (226)	13 [6,33] (1147)		1 [0,3] (305)	6 [3,14] (5551)	

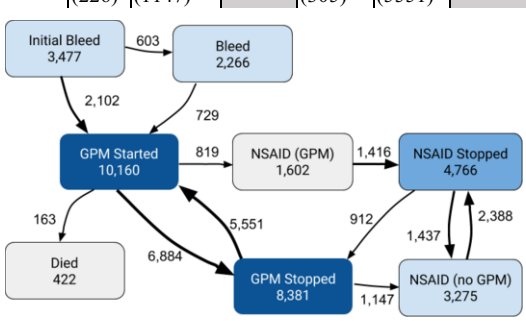

Figure 2– Process Diagram for Indicator I2. NSAID – non-steroidal anti-inflammatory drug, GPM – gastroprotective medication.

Similar results were found in indicator I2 when out of 6,368 NSAIDs, 25% (1,602) were for patients with a pre-existing GPM, 14% (861) were followed almost instantly by a GPM, and 38% (2,388) were for a short-term prescription without a GPM. The results for indicator I3 suggest that GPMs are more frequently co-prescribed with APs with only 22% of AP start events (513 out of 2309), neither having a pre-existing GPM or immediately followed by a GPM.

The event most likely to precede a GiB or peptic ulceration is a previous GiB or peptic ulceration. This is to be expected as it is known that a strong predictor of gastro-intestinal adverse events is a previous bleed or ulceration.

Table 5– Median Duration in Months of Transitions between Events for Indicator I3. IQR in [square] brackets. Number of transitions in (round) brackets. AP – antiplatelet, GPM – gastro-protective medication.

Next event / Event	Bleed	AP (no GPM)	AP (GPM)	AP Stopped	GPM Started	GPM Stopped
Initial Bleed	3 [0,77] (621)	78 [1,287] (390)			4 [1,120] (2197)	
Bleed	2 [1,7] (758)	39 [2,122] (81)	12 [4,54] (52)	11 [2,40] (59)	1 [0,35] (761)	5 [1,24] (286)
AP (no GPM)	7 [1,36] (33)			2 [1,7] (451)	0 [0,6] (851)	
AP (GPM)	7 [2,24] (51)			5 [1,25] (482)		8 [2,31] (160)
AP Stopped	7 [1,23] (30)	6 [2,13] (326)	6 [2,14] (353)		8 [3,29] (181)	0 [0,6] (480)
GPM Started	5 [1,25] (525)		14 [2,47] (540)	7 [2,26] (445)		2 [1,7] (7337)
GPM Stopped	10 [3,33] (246)	10 [2,31] (566)		4 [1,21] (144)	7 [4,18] (6165)	

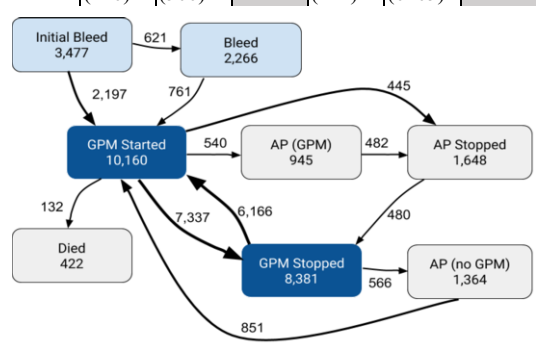

Figure 3– Process Diagram for Indicator I3. AP – antiplatelet. GPM – gastro-protective medication.

The median duration of NSAID and GPM prescriptions is 1 month, suggesting that these medications are typically short-term. Antiplatelets are prescribed at a median length of 5 months and 2 months for patients with and without a pre-existing GPM respectively, suggesting longer term prescriptions.

Discussion

Summary of findings and comparison to existing literature

Little published work on process mining in primary care exists [2]: Dagliati et al. [17] used primary care data to investigate care pathways related to cardiovascular risk of Type II diabetes patients. However, the majority of their data was obtained from secondary care. Another paper used primary care data, but didn't report any results [18]. A further 4 papers used insurance data [19–22] which probably included primary care data, but also secondary care and tertiary care data. Also, the level of data included in insurance datasets is different to that which is routinely collected in primary care for the provision of care. To the best of our knowledge, the process mining performed for this paper is the first performed exclusively using primary care electronic health data.

A GPM was more likely to be co-prescribed to patients receiving a course of antiplatelets than it was to those receiving NSAIDs. The difference in prescription lengths is one possible explanation. When a clinician prescribes a short-term NSAID course, perhaps in response to an acute injury or minor illness, he/she may decide the risk is small enough that co-prescription of a GPM is unnecessary. However, when prescribing a longer-term course, the risk is increased. This might also be true for those on longer courses of NSAIDs for chronic pain conditions. Stratifying medications depending on whether they are short or long term might give further insight into clinicians' behaviour.

GPMs such as proton pump inhibitors can be prescribed for a variety of reasons. For treating an active bleed, to manage the symptoms of gastrointestinal irritation of reflux, or prophylactically for patients at high risk of a bleed – especially when increased because of other medications. Attempting to stratify the GPM events accordingly could again lead to more understanding of prescribing behaviour.

Implications for practice and research

In order to achieve our results, there were several challenges that needed to be overcome which could explain why process mining in secondary care is more prevalent.

Data quality

The quality of healthcare data is limited for many reasons. Events can be incorrectly recorded, unrecorded, or uncoded. All of which limit the confidence and utility of any results generated. Researchers must try and understand the limitations in their data to make best use of it. Primary care data can be thought of as snapshots of coded information that are generated on every contact with the health system. This is different for inpatient secondary care where the entire duration of treatment can be observed and recorded.

While many events recorded in a primary care system may have uncertain veracity, the generation of a prescription is an event we can mine with confidence because, in the UK, virtually all prescriptions are electronically generated in primary care. This is not true for the adverse event of bleeding which may occur in secondary care and may not be coded in the primary care record. To mitigate against this, linked primary and secondary data would be required, and is another reason why process mining exclusively in primary care is not done. Future work should focus on pathways that occur almost exclusively in primary care such as the diagnosis, monitoring and treatment of certain chronic conditions such as hypertension.

Start and end points

Within secondary care, the start of a process can clearly be defined as the admission to hospital, while the end of the process is discharge or death. A patient visiting hospital more than once can be treated as two separate pathways. Within primary care, processes are often cyclical and entangled with other processes. Taking indicator I2 as an example, should the start event be the first instance of peptic ulceration, or should it be the first prescription of an NSAID in a patient with previous peptic ulceration? The former means that each patient only has one pathway, but with potentially multiple cycles, while the latter separates each NSAID prescription into a separate process but then doesn't take the patient's history into account. Detailed consideration must be given to determine whether the primary care processes under analysis have well defined end points, and queries structured to separate data into these individual processes.

Event granularity

The coded events within primary care are not necessarily the events that should make up the event log on which process mining depends. An example is medications where the patient record contains the prescription events, whereas the events of interest on a care pathway would be when the clinician has started, stopped, or altered the dosage of a medication. This is also true for measurements, where a series of blood pressure (BP) values do not necessarily constitute events, but the occurrence of two systolic BPs >140 mmHg within two weeks might be a trigger to investigate a diagnosis of hypertension and could therefore be considered an event to process mine. However, this introduces a bias as the researcher must decide *a priori* what constitutes an event. Is it that a BP was taken, that the BP was high or that some combination of values was measured over a certain period of time? Careful consideration must be made to convert the raw data into an event log, but this is not straightforward and is largely subjective.

Medications

The lack of a stop event for medications requires an extra processing step to determine when a patient's medication has expired. There is also no way of knowing whether a medication once collected is in fact consumed by the patient, or if the patient is using over the counter medications. This is less of an issue in inpatient secondary care when the both the prescription and administration of medication can be monitored and recorded.

Memory

The process mining diagrams that we have produced are heuristic nets which are memory-less: each transition in the process map is taken in isolation without consideration of prior events. By redefining the start events of NSAID and antiplatelets to take into account whether a GPM was already prescribed allowed us to introduce an element of memory into the system. This is useful to better understand the various pathways, however future work using other process mining modelling techniques, such as causal nets [23], might produce better results.

Conclusions

Primary care data in the UK has reliably coded prescribing information and process mining can be successfully applied leading to results that may be useful for clinical decision support systems and improving patient safety.

However primary care data presents several unique challenges. Careful pre-processing must first be undertaken, but this is subjective and must therefore be sensibly performed and meticulously recorded in order to facilitate scrutiny and reproducibility. The use of a clinical reference group to review and confirm data quality and provide insight into the direction of research would be beneficial. Other more powerful process mining and machine learning techniques could be applied now that the initial problems with primary care data have been considered and to some extent addressed.

Acknowledgements

This article was funded by the National Institute for Health Research Greater Manchester Patient Safety Translational Research Centre (NIHR Greater Manchester PSTRC). The views expressed are those of the author and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

References

[1] W.M.P. van der Aalst, Process Mining: Discovery, Con–formance and Enhancement of Business Processes, Springer, 2011. doi:10.1007/978-3-642-19345-3.

[2] R. Williams, E. Rojas, N. Peek, and O.A. Johnson, Process mining in primary care: A literature review, 2018. doi:10.3233/978-1-61499-852-5-376.

[3] A.J. Avery, M. Ghaleb, N. Barber, B.D. Franklin, S.J. Armstrong, B. Serumaga, S. Dhillon, A. Freyer, R. Howard, O. Talabi, and R.L. Mehta, The prevalence and nature of prescribing and monitoring errors in English general practice: A retrospective case note review, *Br. J. Gen. Pract.* **63** (2013) e543-53. doi:10.3399/bjgp13X670679.

[4] R.A. Elliott, E. Camacho, F. Campbell, D. Jankovic, M. Martyn St James, E. Kaltenthaler, R. Wong, M.J. Sculpher, R. Faria, and M. Rita Faria, Prevalence and Economic Burden of Medication Errors in the NHS in England: Rapid evidence synthesis and economic analysis of the prevalence and burden of medication error in the UK, *Policy Res. Unit Econ. Eval. Heal. Care Interv.* (2018).

[5] Ekaterini Blaveri, Mapping the Landscape of UK Health Data Research & Innovation A snapshot of activity in 2017, 2017. https://mrc.ukri.org/documents/pdf/mapping-the-landscape-of-uk-health-data-research-and-innovation-report/ (accessed November 13, 2018).

[6] A.R. Tate, S. Dungey, S. Glew, N. Beloff, R. Williams, and T. Williams, Quality of recording of diabetes in the UK: How does the GP's method of coding clinical data affect incidence estimates? Cross-sectional study using the CPRD database, *BMJ Open*. **7** (2017) e012905. doi:10.1136/bmjopen-2016-012905.

[7] C.W. Günther, and W.M.P. van der Aalst, Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics, in: 2007: pp. 328–343. doi:10.1007/978-3-540-75183-0_24.

[8] R. Spencer, B. Bell, A.J. Avery, G. Gookey, and S.M. Campbell, Identification of an updated set of prescribing-safety indicators for GPs, *Br. J. Gen. Pract.* **64** (2014) e181-90. doi:10.3399/bjgp14X677806.

[9] A.J. Avery, S. Rodgers, J.A. Cantrill, S. Armstrong, K. Cresswell, M. Eden, R.A. Elliott, R. Howard, D. Kendrick, C.J. Morris, R.J. Prescott, G. Swanwick, M. Franklin, K. Putman, M. Boyd, and A. Sheikh, A pharmacist-led information technology intervention for medication errors (PINCER): A multicentre, cluster randomised, controlled trial and cost-eff ectiveness analysis, *Lancet*. **379** (2012) 1310–1319. doi:10.1016/S0140-6736(11)61817-5.

[10] R. Williams, R. Keers, W.T. Gude, M. Jeffries, C. Davies, B. Brown, E. Kontopantelis, A.J. Avery, D.M. Ashcroft, and N. Peek, SMASH! The Salford medication safety dashboard, *J. Innov. Heal. Informatics*. **25** (2018) 183. doi:10.14236/jhi.v25i3.1015.

[11] O. Corporation, OptimizeRx, (n.d.). https://www.optimizerx.com/ (accessed November 13, 2018).

[12] R. Williams, B. Brown, N. Peek, and I. Buchan, Making Medication Data Meaningful: Illustrated with Hypertension, *Stud. Health Technol. Inform.* **228** (2016) 247–51. doi:10.3233/978-1-61499-678-1-247.

[13] R. Williams, E. Kontopantelis, I. Buchan, and N. Peek, Clinical code set engineering for reusing EHR data for research: A review, *J. Biomed. Inform.* **70** (2017) 1–13. doi:10.1016/j.jbi.2017.04.010.

[14] Node.js, (n.d.). https://nodejs.org/en/ (accessed December 19, 2017).

[15] Fluxicon, Process Mining and Automated Process Discovery Software for Professionals - Fluxicon Disco., (n.d.). https://fluxicon.com/disco/ (accessed October 31, 2018).

[16] Ministry of Housing Communities & Local Government, English indices of deprivation 2015, 2015. doi:10.1016/j.diamond.2017.05.012.

[17] A. Dagliati, L. Sacchi, C. Cerra, P. Leporati, P. De Cata, L. Chiovato, J.H. Holmes, and R. Bellazzi, Temporal data mining and process mining techniques to identify cardiovascular risk-associated clinical pathways in Type 2 diabetes patients, in: IEEE-EMBS Int. Conf. Biomed. Heal. Informatics, IEEE, 2014: pp. 240–243. doi:10.1109/BHI.2014.6864348.

[18] T. Meneu, V. Traver, S. Guillen, B. Valdivieso, J. Benedi, and C. Fernandez-Llatas, Heart Cycle: Facilitating the deployment of advanced care processes, *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*. (2013) 6996–6999. doi:10.1109/EMBC.2013.6611168.

[19] R.P. Hilton, N. Serban, and R.Y. Zheng, Uncovering Longitudinal Healthcare Utilization from Patient-Level Medical Claims Data, (2016).

[20] K. Tóth, I. Kósa, A. Vathy-Fogarassy, and D. Hayn, Frequent Treatment Sequence Mining from Medical Databases, *Stud. Health Technol. Inform.* **0** (2017) 211–218. doi:10.3233/978-1-61499-759-7-211.

[21] J. Lismont, A.S. Janssens, I. Odnoletkova, S. vanden Broucke, F. Caron, and J. Vanthienen, A guide for the application of analytics on healthcare processes: A dynamic view on patient pathways, *Comput. Biol. Med.* **77** (2016) 125–134. doi:10.1016/j.compbiomed.2016.08.007.

[22] G.G.T. Lakshmanan, S. Rozsnyai, and F. Wang, Investigating clinical care pathways correlated with outcomes, *Bus. Process Manag.* (2013) 323–338. doi:10.1007/978-3-642-40176-3_27.

[23] W. Van Der Aalst, A. Adriansyah, and B. Van Dongen, Causal nets: A modeling language tailored towards process discovery, in: Lect. Notes Comput. Sci., Springer, Berlin, Heidelberg, 2011: pp. 28–42. doi:10.1007/978-3-642-23217-6_3.

Address for correspondence

Richard Williams - richard.williams@manchester.ac.uk

G303a, JMF Building, University of Manchester, M13 9PL, UK