

# UNIVERSITY OF BIRMINGHAM

## Research at Birmingham

### A process mining and text analysis approach to analyse the extent of polypharmacy in medical prescribing

Weber, Philip; Palmer, Ruth; Litchfield, Ian; Lee, Mark

DOI:

[10.1109/ICHI.2018.00008](https://doi.org/10.1109/ICHI.2018.00008)

License:

None: All rights reserved

*Document Version*

Peer reviewed version

*Citation for published version (Harvard):*

Weber, P, Backman, R, Litchfield, I & Lee, M 2018, A process mining and text analysis approach to analyse the extent of polypharmacy in medical prescribing. in The Sixth IEEE International Conference on Healthcare Informatics (ICHI 2018). IEEE Xplore, The Sixth IEEE International Conference on Healthcare Informatics , New York City, United States, 4/06/18. <https://doi.org/10.1109/ICHI.2018.00008>

[Link to publication on Research at Birmingham portal](#)

**Publisher Rights Statement:**

Checked for eligibility: 02/05/2018

Published in IEEE Xplore: 10.1109/ICHI.2018.00008

**General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

**Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# A Process Mining and Text Analysis Approach to Analyse the Extent of Polypharmacy in Medical Prescribing

Philip Weber\*, Ruth Backman†, Ian Litchfield† and Mark Lee\*

\*School of Computer Science, University of Birmingham, Edgbaston, Birmingham, United Kingdom  
Email: dr.philip.weber@ieee.org, m.g.lee@cs.bham.ac.uk

†Institute of Applied Health Research, University of Birmingham, Edgbaston, Birmingham, United Kingdom  
Email: r.backman@bham.ac.uk, i.litchfield@bham.ac.uk

**Abstract**—Prescription of conflicting concurrent medications, polypharmacy, is recognised as a significant problem in the UK, but its extent is not fully known. We combined a process mining approach with text analytics, to discover prescription processes for patients from five primary care sites in the UK West Midlands. Free-text prescription instructions were combined with online knowledge about drug interactions to reveal that almost 62% of patients were prescribed with medication with some level of interaction, during a two year period. We describe a novel domain-specific approach to reduce the complexity of the mined processes, which will nevertheless be applicable to other flexible environments such as knowledge work. We also highlight difficulties encountered in accessing, interpreting and processing the data, which may be significantly mitigated through wider adoption of a mindset of curating data for automated analysis.

**Index Terms**—Process mining, text analysis, polypharmacy, health informatics, big data, web scraping.

## I. INTRODUCTION

Polypharmacy refers to the concurrent prescription of multiple medications for a single patient [18], [22]. It is a growing problem [5], in particular due to increasing numbers of patients suffering from multiple chronic diseases (multimorbidity), arising from factors including ageing populations and improved treatment of chronic disease [46]. The increased burden of treatment is a particular concern for those with multimorbidity as they frequently take many medications related to each individual condition. Polypharmacy has the potential to reduce treatment adherence [34], [58], and is associated with negative health outcomes such as cognitive impairment [31] or hospital stays due to adverse drug events [36]. These negative outcomes are most evident amongst the elderly [57], [44] where they are compounded by the drug-drug interactions found when three or more sets of clinical guidance are combined [14]. When not carefully monitored and regularly reviewed, polypharmacy can result in adverse interactions between a patient’s medications, reducing their effectiveness or causing additional problems. These problems are recognised in the UK, but their extent is not fully known.

Prescribing occurs as part of following the care pathway (or process) for treating a patient for a particular condition, and polypharmacy is just one of the problems that can occur in

following several such pathways concurrently. Understanding these processes would give insight into the root causes leading to problems and enable effective improvements to be put into place. We are formally modelling care pathways using process notations [60], in order to detect and resolve conflicts including between medication, lifestyle recommendations and scheduling. Such resolutions have the potential to improve the effectiveness and efficiency of pathways, reduce the cost and patient burden from unnecessary medication and resultant complications, and improve patient experience.

Process mining [51] enables the discovery and analysis of processes, such as those found in service industries [48] or healthcare [32], [41], from data. We describe a pilot study using process mining principles and text analytics to analyse the drug prescription aspect of care processes, using prescription data from primary care sites in the UK West Midlands region, to gain insight into the prevalence of polypharmacy. In this study we focus only on patients with six common chronic conditions<sup>1</sup>; there will be a much greater extent of interactions and polypharmacy across UK primary care as a whole. We ask two main questions: firstly, which conflicting drugs are prescribed together, how often, and how severe are any potential interactions? Secondly, what frequently-occurring patterns of prescription can be identified to give insight into the underlying causes of polypharmacy and enable mitigations to be developed? More details of the clinical rationale can be found in the study protocol [5].

Process mining was developed to discover and analyse a business’ processes using ‘event data’ routinely recorded by its information systems. It has been successfully applied in healthcare [32], [41], e.g. to model hospital workflows [38]. While many graph mining methods exist for learning patterns of connections and causal relationships between entities [56], process mining can account for activities occurring in parallel (concurrency) and with non-zero duration, both of which are crucial for our analysis. We conduct our analysis with a process mining mindset, considering the medication that a

<sup>1</sup>Chronic Obstructive Pulmonary Disease (COPD), Coronary Heart Disease (CHD), Hypertension, Osteoarthritis, Type 2 Diabetes and Depression.

patient takes as an ‘activity’ occurring over a period of time, prescriptions as ‘events’ indicating the start of an activity, and asking questions about the amount and type of concurrency and how the pattern of prescriptions (process) evolves over time. Mining complex, noisy and low-structured processes is an active area of process mining research, to which this work may contribute.

In Section II we introduce process mining and how it applies to the analysis of prescription data. We then describe (Section III) our data analysis process to investigate the polypharmacy question. We focus on text processing to transform unstructured clinical records into suitable data for process mining; a process mining, generalisation and visualisation approach appropriate for the medication prescription data; and obtaining initial results on polypharmacy (Section IV). We found that almost 62% of patients in our sample received medication with some level of interaction, which our clinical study [5] will explore in more detail. We provide initial analysis in Section V, and briefly discuss some of the problems we encountered in obtaining and processing the data, which could be significantly mitigated by relatively simple changes and adopting a mindset of curating data for the purposes of automated analysis. The final section (VII) outlines future research to integrate the analysis into the wider question of discovering and resolving conflicts between care pathways, and to extend the process mining ideas to flexible and noisy processes more generally.

## II. PROCESS MINING APPROACH

Process mining [51] uses ‘event’ data routinely collated by an organisation’s computer systems to learn models of its business processes, and to represent and reason about them both visually and formally. It is related to data mining but distinguished by considering the process as a whole. A similar methodological approach is often needed, i.e. select and pre-process data, carry out mining, and interpret the results [40].

Typical data mining techniques seek to identify relationships between variables statically (association rule mining), over time series (e.g. trends), or find patterns over sequences of measurements (sequential or structural pattern mining). Process mining in contrast explicitly considers whole processes and the activities involved, especially the relationships (causal, concurrent, cyclical or mutually exclusive) between them. A recent hybrid approach, Local Process Mining [47], applies similar methods to discover frequent process fragments, conceptually closer to sequence mining. Further aspects include comparing processes, modelling resource usage or performance, or discovering logical inconsistencies in the process. Whereas the goal of data mining is often predictions of the future states or trends, process mining initially focusses on modelling and visualising the current state of the process. Data mining concepts may then be employed alongside human interpretation, for subsequent processing such as clustering or prediction of process outcomes.

Defining criteria of process mining are the expectation that activities may take time, be concurrent (overlap in time) [53], data and processes may be complex or ‘noisy’ [21],

[61], and may be of interest from several viewpoints (e.g. causal relationships, organisational structures, or ‘resources’). Since we are interested in the interaction between concurrent medication, defined by the events of prescription, and patterns of prescription that may lead to such interactions, process mining seems an appropriate approach to this analysis.

A process mining algorithm requires a minimum of three pieces of information to be recorded for each event:

- (1) An *activity ID* specifies what activity took place, which might be receipt of an invoice or repair made. We use the name of the medication prescribed.
- (2) A *case ID* links related activities, e.g. for a particular invoice or fault call. Patient IDs serve this purpose in our analysis since we are interested in relations between all medications prescribed to each patient.
- (3) A timestamp specifies when the activity occurred, ideally at a high level of granularity. We have the date the prescription was issued.

From ‘traces’ (sequences) of these events a process discovery algorithm attempts to infer so-called ‘dependency relations’ between activities, and thus a process model showing the patterns in which activities can occur, i.e. in sequence, parallel, mutually exclusive, or cyclically. Models have been represented in various languages such as Petri nets [54], but the use of Business Process Model and Notation (BPMN) [37] is becoming a *de facto* standard. Following discovery, models can be analysed for conformance to (similarity with) other processes or business rules [43], [3], [13], performance or bottlenecks [2], [52], interaction between resources [7], or to predict outcomes or simulate process changes [55], [39], [27].

Although many process mining techniques require these three pieces of data for an event, other approaches try to determine a process from less [16], while many use additional data to either improve discovery, e.g. using both activity start and end times [9], [28], or to ‘enrich’ the process model. For example a clinical process might be enhanced with clinician names, results of tests or GP comments, especially any comments relating to shared decision making with the patient.

Process mining has been successfully used in secondary care [41], [32], [38], but to the best of our knowledge not to investigate patterns of medical prescription. We view the sequence of prescribing medications to patients as a type of process. A typical business process, for example for processing an invoice, has a defined start and end point (receipt and payment of the invoice), and a model of the process defines the possible sequences of activities. The steps taken for a particular invoice describe one process ‘case’. Our situation differs in that the start and end point of each case (a single patient) is artificially imposed by the data extract. However, we face similar challenges: concurrency, ‘noisy’ and diverse data (many medications and feasible combinations), and complex models within which we want to find the most frequent behaviour. We consider each prescription of a medication as an ‘activity’ taking place for a given time, associated with start (issuing the prescription) and end (completing the course)

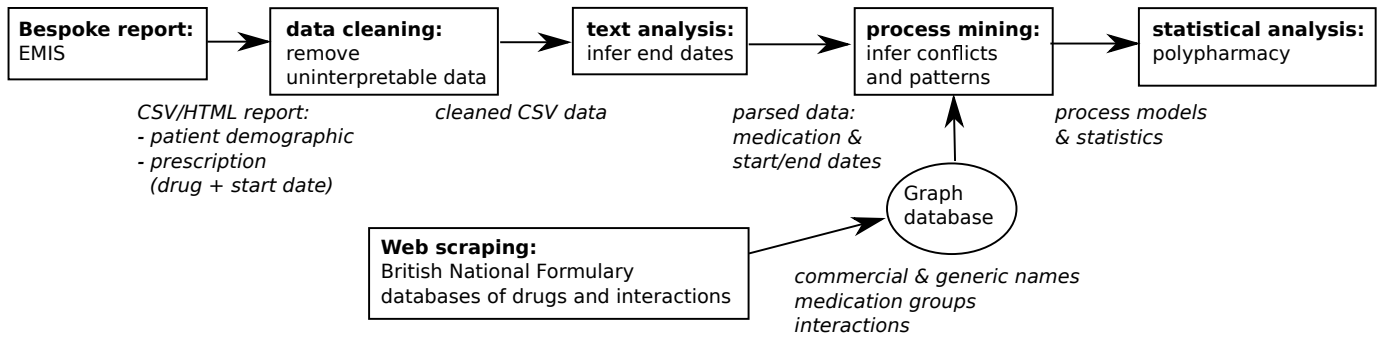


Fig. 1: Data processing pipeline for analysis of prescription data for polypharmacy.

events. Different medication may be prescribed sequentially, concurrently (polypharmacy) or as alternative paths through the process (e.g. for treating different conditions).

A particular problem faced in process mining in healthcare contexts is that processes tend to be complex and dynamic, involve multiple disciplines or departments, and may be ad-hoc [40], [29], allowing many variants. Process discovery then results in so-called ‘spaghetti’ models [50], [15] – too complex to interpret or visualise. Numerous approaches to manage such models have been proposed, from focussing on the most-followed activities and paths [21], [28], to pre-clustering process traces [40], [17], [6], to semantic interpretation of activity names [11], [35]. It is, however, still an open problem, with no universally-applicable solutions. The most appropriate approaches may be data- or application-dependent. We propose a representation appropriate for the prescription data, which represents concurrency (between multiple medications) implicitly.

### III. DATA ANALYSIS

In the next sections we describe the steps taken to obtain, clean and process prescriptions and interactions data into a form suitable for the analysis. The process is summarised in Fig. 1. We broadly follow what is becoming the standard approach to process mining projects; obtain and pre-process log files, carry out control-flow process discovery, further analysis, and report results [8], [40]<sup>2</sup>.

#### A. Data specification and extraction

Five primary care sites (General Practitioner (GP) surgeries) were selected, having a variety of demographic characteristics (Table I). Each site has a PC running Windows XP and the Egton Medical Information Systems (EMIS) Web<sup>3</sup> application and database providing access to real-time patient data and care process management. A custom report<sup>4</sup> was used to extract prescription data covering a period of just over 2 years (27 months, from 2 Dec 2014 to 28 Feb 2017), for patients diagnosed with at least one of the six major chronic diseases chosen. Data were extracted for between 515 and 2,940

Site	Raw Event Log		Processed Data	
	Patients	Prescriptions	Patients	Prescriptions
Bo	515	36,940	507	29,064
Ma	1,259	158,953	1,238	126,851
QB	1,343	125,441	1,339	107,786
EH	2,541	268,696	2,502	215,243
KN	2,940	208,557	2,895	160,694
<b>Total</b>	<b>8,598</b>	<b>798,587</b>	<b>8,481</b>	<b>639,638</b>

TABLE I: Prescription data extracted from five sites, showing patient counts and number of prescriptions in the source data, and after pre-processing to extract prescription end dates where feasible.

patients per site (8,598 in total), 36,940–268,696 prescriptions (total 798,587). Ethical approval was granted for this work by the University of Birmingham Research Ethics Committee, with reference number ERN\_16-0074.

Each record details the issuing of a prescription for a single medication to a patient. In four cases the extract was formatted as an Excel spreadsheet; in the final case the PC was underpowered and the extract was made to HTML and post-processed. The final output was converted to comma-separated (CSV) format for subsequent processing. The fields extracted are listed in Table II. The data includes pseudonymised patient identifiers, demographics (gender and age at date of extract), date of issuing the prescription, structured text specifying the medication, quantity, and free text from the GP to the patient specifying the amount of medication and prescription regimen. Data were also extracted to enable future analysis by the demographics of the practice and by medical condition, but are not the subject of the analysis described in this paper.

#### B. Determine Prescription End Dates

To assess polypharmacy, we require knowledge of when prescriptions were made, and either how long a medication was being taken by a patient, or prescription end dates. The latter can be inferred in many cases from four key fields listed in Table II: (1) Name, Dose and Quantity (NDQ), (2) Dose, (3) Quantity, (4) Quantity Unit.

‘NDQ’ is relatively structured and includes medication name and information such as tablet size (e.g. ‘Metformin 850mg tablets’, ‘Aciclovir 5% cream’).

<sup>2</sup>Code at <https://bitbucket.org/uobmitcon/ichi2018/>.

<sup>3</sup><https://www.emishealth.com/products/emis-web/>.

<sup>4</sup>Report and data extract by PRIMIS <http://www.nottingham.ac.uk/primis/>.



Field(s)	Example Fields	Description
Demographics	QB, e777ab. . . , Male, 67, EH, 945ba6. . . , Female, 51,	Site code, anonymised patient ID, gender and age (18–104).
Date of Issue	11-Feb-15	Date the prescription was issued to the patient.
Name, Dosage and Quantity (NDQ)	'Metformin 850mg tablets' 'Zerobase 11% cream (Thornton & Ross Ltd)'	Medication name and details (structured).
Dose	'Take one three times a day for a week' 'To be taken as directed'	GP instructions to the patient (unstructured).
Quantity	168	Number of tablets supplied, volume of liquid, etc.
Quantity Unit	tablet	Tablet, gram, unit, etc.

EH, 6ce8921. . . , Female, 68, 03-Aug-16, 'Metformin 850mg tablets', 'Two Tablets Daily At The Same Time Each Day', 56, tablet  
EH, cbac5a6. . . , Male, 75, 30-Dec-16, 'Zerobase 11% cream (Thornton & Ross Ltd)', 'use as moisturiser as often as necessary', 500, gram

TABLE II: Summary of the data fields extracted, and two example records.

'Dose' is the critical field, containing free-text notes from the GP specifying how much medication is prescribed, when and how it should be taken, and how often.

'Quantity' indicates how much medication was prescribed (number of tablets or volume), and

'Quantity Unit' specifies the method of delivery from about 30 options including 'tablet', 'ml', 'spray', 'syringe'.

Although 'Dose' is free text, there is some limited structure, perhaps due to the constrained domain (prescribing) and the EMIS Web application remembering previously-entered data. Thus we can construct a grammar to parse many of the entries for the required information.

We infer prescription end dates using a sequence of data cleaning and parsing as follows, to extract the quantity of medication to be taken per dose, and frequency, from which to infer the number of days the medication will last, and thus the end date.

1) *Data Cleaning (Pre-processing)*: Before parsing, data cleaning was necessary due to the unstructured format of the 'Dose' field in particular, data missing by design or inappropriate for automated processing, and for simplification for further processing:

- (1) Free-text entry leads to typographical errors and inconsistent use of grammar such as plurals. These were identified and corrected. We also removed text redundant for the task at hand, such as GP instructions (e.g. 'dissolve the contents of', 'for subcutaneous injection').
- (2) In almost 9% of cases it is impossible to determine the amount of medication used, and hence the end date: medication specified with a non-specific dose ('emollient', 'shampoo' and similar); 'Quantity Unit' such as 'device' or 'spray' without specification of the size of the device or amount of medication delivered; and prescriptions with vague instructions, especially all variations of 'use as directed'. These were all excluded, together with records with data errors such as unexpectedly empty fields, or basic data problems such as non-numeric quantities.

These records were excluded from subsequent processing.

- (3) Text such as 'into both eyes', 'in each nostril' was used to determine a multiplier to simplify the final dose calculation.

2) *Text Processing to Obtain End Dates*: A grammar was developed to parse the unstructured 'Dose' to extract quantity and time information<sup>5</sup>. The basic form of the field is

Dose ::- <directive> <quantity> <unit>  
<frequency> <time-spec>

as seen in the examples in Table II. These main elements can however be combined in the data in many ways to convey similar information ('take one tablet daily' vs 'one tablet to be taken each morning'), some may be omitted ('1 tablet mane') or other information included ('one tablet daily with food'). We combine several sub-grammars for processing broad categories of medication which differ in general format, e.g. according to prescription by volume (liquids), by discrete unit (e.g. tablets) or other (syringes) as identified by the 'Quantity Unit' field.

In outline the parser recognises numeric values as digits, fractions, words or ranges; time specifications given numerically, using clinical abbreviations ('o.d.', 'mane') or colloquially ('every morning', 'with evening meal'); and amounts given numerically or approximately ('half a mil', '1 teaspoon'). These are converted to a numeric quantity of medication per dose (e.g. number of tablets or volume of liquid) and frequency, from which a nominal amount of medication, possibly fractional, per day is calculated. This is combined with the 'Quantity' field to give the duration of the prescription and hence the end date. The more structured 'NDQ' field provides additional information, e.g. the volume per dose of medication administered by syringe.

In the first example in Table II, 56 tablets were prescribed on 3 Aug 2016, 'Two Tablets Daily At The Same Time Each

<sup>5</sup>Grammar implemented using the Python `pyarsing` module.

Day’, so 2 tablets per day last for 28 days, thus ending on 31 Aug 2016. No end date can however be inferred from the second example, since it is not known how much moisturiser will be consumed per use, nor how often will be necessary, so this record is excluded.

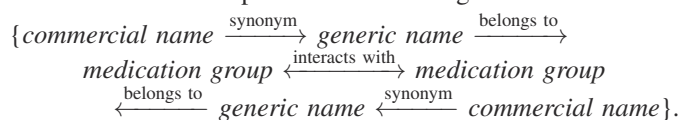
3) *Remarks on the Learning Process:* Since this is a real-world dataset not created with automated processing in mind, there is no ground truth which we can use to quantify the precision or recall of the grammar against known training or test corpora. Instead, qualitative assessment by experts was used to iteratively refine the grammar to identify the main text structures, test, refine to account for un-parsed records, repeating until the returns in reduced exclusion rate no longer justified the effort. (Future work could perhaps attempt to create simulated data sets faithful to the observed data, but these are likely to be overly noise-free, leading to artificially positive evaluations of learning accuracy.)

We made various assumptions in interpreting the data, principally to err in the direction of over-reporting polypharmacy rather than risk omitting dangerous combinations of drugs. Unclear dose or time specifications were simplified where possible to the value leading to the longest period of prescription (e.g. ‘one or two’ tablets becomes ‘one’). Similarly, complex specifications such as ‘every night for 2 weeks then twice a week for 3 months’ were reduced to the first time specification (assuming the rate of prescription reduces). Fixed-term time specifications (actual dates) were ignored, and alternating dose specifications such as ‘20mg/40mg alternate days’ were simplified to the first. These and other limitations, including ‘outlier’ text structures outside the scope of the grammar, could be resolved with additional development effort. Just under 20% of the extracted records were removed as not processable by the pre-processing (9%) and grammar (11%).

### C. Inferring Medication Conflicts

Medication conflicts are sourced from an online presentation of the British National Formulary (BNF)<sup>6</sup>. The BNF lists details of medications, and ‘standard’ and ‘strong’ interactions between individual and groups of medications<sup>7</sup>. We store the interacting drugs as nodes in a graph database, creating two relations: standard and strong interactions. Further relations define groups of drugs, to record interactions inherited from the group, and to map commercial to generic drug names. This mapping is found at a different section of the website and is necessary because the prescription data refers to drugs using a mix of both generic and commercial names.

Graph databases provide appropriate mechanisms to record a chain of relationships between two drugs:



<sup>6</sup>BNF (<https://www.bnf.org/>) data was ‘scraped’ from <https://bnf.nice.org.uk/> using the Python `lxml.etree` module `html`. Data was stored in a Neo4j (<https://neo4j.com/>) graph database accessed using the `py2neo` module.

<sup>7</sup>The most recent BNF allows for four levels of interaction.

A commercial name (e.g. ‘Nurofen’) may be a synonym for a generic name (‘ibuprofen’) and perhaps a member of a group of drugs (‘non-steroidal anti-inflammatory drugs’) which interact with another group (‘corticosteroids’). A member of the latter group (‘betamethasone valerate’) may also have a commercial instantiation (‘Betnovate’). Medication may be specified as, and interactions require identification between, various of these links, e.g. ‘Nurofen’ prescribed with ‘anti-inflammatory drugs’, or ‘ibuprofen’ with ‘Betnovate’. The database query language enables such chains to be retrieved, stored and queried efficiently so that the interactions can be captured at whichever level they are specified.

We retrieved lists of drugs from two pages of the BNF, firstly associating commercial with generic drug names, secondly linking to the interactions. These lists were inconsistent, necessitating basic text pre-processing: to match plurals, name variants (‘Hepatitis A vaccine’/‘Hepatitis vaccines’) and remove chemical specifiers (‘hydroxide’, ‘acetate’, etc.). We also reduce detailed commercial drug names to their basic format as used in the prescription data, e.g. ‘Alecensa 150mg capsules (Roche Products Ltd)’ becomes ‘Alecensa’).

In total 2,360 drug names (generic and commercial) were extracted, 927 group relationships, 5,770 standard and 5,151 strong interactions, and 1,232 commercial drug names. Querying the database to identify all unique interactions accounting for name synonyms and groups results in 148,567 standard, 61,113 strong interactions.

### D. Process Mining

As described in Section II, we take a process mining approach to interpret the prescriptions data. The ‘activity’ that a patient is prescribed a medication occurs over a time frame (between ‘Date of Issue’ and the inferred end date), and multiple such ‘activities’ may overlap wholly or partially (polypharmacy). Our present interest is to detect such overlaps to analyse the prevalence of drug interactions, but a natural extension is to identify frequent patterns of prescription and understand the underlying causes. Therefore we build ‘process models’ of the patterns of prescription, which may be interpreted formally and visually, and from these obtain the data necessary for later statistical analysis.

1) *Standard Process Mining:* We first carried out an exploratory analysis of the prescription data using existing process mining algorithms (Alpha [53], Heuristics Miner [61] and Inductive Miner [28]) and tools (ProM<sup>8</sup>, Disco<sup>9</sup> and Apromore<sup>10</sup>). Process mining requires as a minimum the first three data fields shown in Table III together with the mappings from our data. We additionally provide the inferred end dates to enable activity durations to be taken into account. We found that the data was too heterogeneous to produce visually useful models, producing classically ‘spaghetti’ visualisations [50], [15]. While there exists much literature on dealing with this problem, including separating multiple processes entangled in

<sup>8</sup><http://www.promtools.org/>.

<sup>9</sup><https://fluxicon.com/products/>.

<sup>10</sup><http://apromore.org/>.

Required Field	Mapped Data Field	Process Mining Description
Case ID	Patient ID	Entity involved, typically invoice or fault number, etc.
Event ID	Drug Name	Steps involved in the process.
Start Time	Date of Issue	Start time of the activity.
End Time	Inferred End Date	Activity end time.

TABLE III: Minimal fields for process mining. We regard a single patient as a ‘case’, and the medications prescribed as the ‘activities’ in the process, indicated by the ‘events’ of their prescription. Note that end times are not required by many process mining algorithms, but are critical for our analysis.

the data (e.g. [20], [40], [17], [6]), and focusing on the most frequent process patterns to produce manageable models [61], [21], [28], these raise questions over interpretation and the validity of the complexity-reducing decisions which are made. Therefore we took a bespoke approach, described next.

2) *Domain-Specific Complexity Reduction*: To restrict the complexity of the models resulting from process mining in a manner which is faithful to the data domain and the questions we investigate, we made two simplifying decisions:

- (1) concentrate on per-patient models, and
- (2) model concurrency implicitly, by merging concurrent medication into single medication groups.

The first decision allows for simpler process models to be produced while still collecting statistics on frequency of medication interactions. The models allow visualisation of basic patterns of prescription, and the method provides a basis for domain-specific generalisation (Section III-D4) and future analysis of patterns of prescription across the whole patient cohort in a principled manner. The second decision is driven by the fact that the entity of interest is *groups of medication* prescribed at any one time, rather than the times when individual drugs are prescribed or completed. Therefore we aggregate concurrency into the nodes in our process maps, representing a unique group of concurrent medication in each node (Fig. 2). Each time a prescription is started or completed, this group changes, and a new node is created and linked appropriately. In this way there is no concurrency between nodes in the model, which we represent as a type of finite state machine and annotate with frequency, duration and interaction information. Per-patient statistics on polypharmacy interaction can be simply extracted from these models.

3) *Domain-Specific Process Mining*: We mine a ‘prescription process’ for each patient:

- (1) Extract and simplify medication names from ‘NDQ’ (remove information on tablet size etc.) to reduce the granularity of activity names.
- (2) Sort ‘Date of Issue’ and ‘End Date’ from each record into an ordered list of start/end prescription events. On clinician advice, we assume that no prescription lasts less than one day, multiple contiguous prescriptions of the same medication (re-issues) should be merged, and mul-

tiply simultaneous prescriptions (e.g. multiple boxes of tablets) of the same drug should be treated as sequential.

- (3) Build a Probabilistic Prefix Tree Automaton (PPTA) [10] from this consolidated sequence of events. For a single patient, this is simply a sequence of nodes representing groups of concurrent drugs (Fig. 2(a), connected by arcs when the group changes. Adding multiple patients into the same structure produces a tree labelled with frequency of change between the nodes.
- (4) Naïvely construct a Probabilistic Deterministic Finite Automaton (PDFA) [10], [59] from the PPTA, where no state (group of medications) is repeated. The PPTA is parsed depth-first from the start node, noting the groups of medications represented by each node. When a node  $n$  is encountered representing a group  $M$  of medications already seen in node  $n'$ , the target of its unique input arc ( $n$ ) and source of any output arcs are connected instead to  $n'$ , resulting in a reduced, connected, model with cycles (Fig. 2(b)). Depending on the pattern of prescription for a patient, the PDFA may be little different from the PPTA (e.g. very varied groups of concurrent medications), or much more similar to a traditional process model, with a defined start and end node, showing a pattern of regular prescription (Fig. 2(b)). Although the PDFA is labelled from the data with frequencies with which each node and arc is passed, treating it as probabilistic allows for future generalisation and analysis using methods from the automata and regular language learning literature (e.g. [10], [23]), e.g. for frequent patterns of prescription.

The resulting models are annotated for standard and strong interactions, and duration of prescription, from which we obtain statistics for analysing polypharmacy, summarised in Table IV. Fig. 2(b) shows a relatively compact example of a ‘prescription process’ for a single patient, consolidated from a long sequence of prescription combinations (Fig. 2(a)), suggesting a relatively structured prescription process. Standard (orange) and strong (red) interactions are highlighted. The process is entered at the top, at the start of the time period for which the data was collected, and exited at the double-bordered (green) box which shows the prescription at the end of the time period. The weights of nodes and arcs give some indication of how much time was spent in each node, and how often each arc was traversed. The dotted box indicates that there were frequently gaps when no medication was prescribed – although this may reflect data problems such as mismatch between the date a prescription was issued and actually taken.

4) *Domain-Specific Generalisation and Analysis*: We apply methods to generalise, or reduce the complexity of the model, in a manner which is sympathetic to the domain. At present these methods are purely heuristic. Less significant nodes (short-duration and small groups of medication) are absorbed into their neighbours (Fig. 3(a)), according to some thresholds, then strongly-overlapping neighbours (parent-child and sibling nodes) merged (Fig. 3(b)). Iteratively, this will reduce a complex model to a very simple one – ultimately

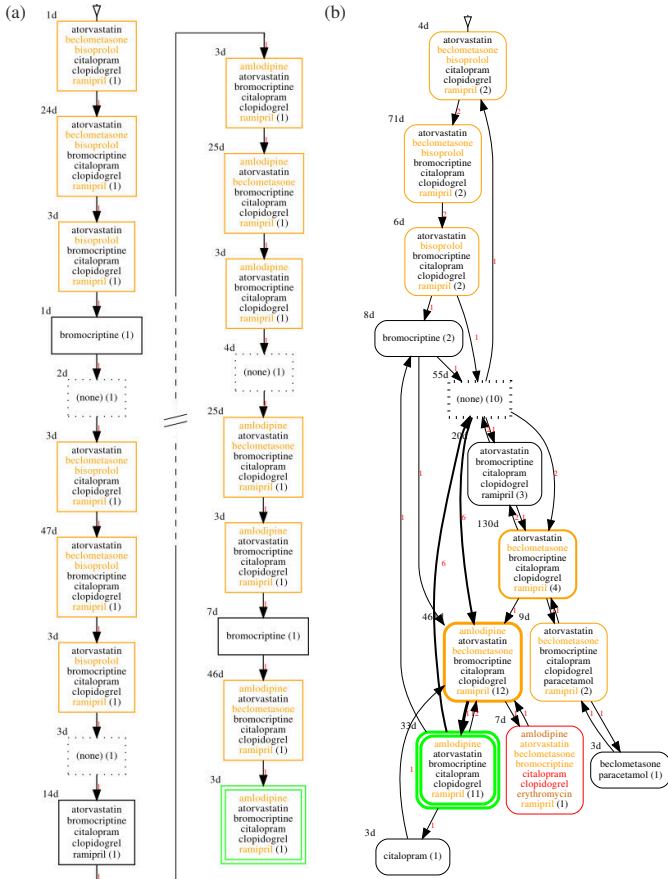


Fig. 2: (a) Sequence of prescription events for a single patient, (b) mined ‘prescription process’ PDFA. Nodes represent a group of concurrently-prescribed medications, annotated with the number of times the combination occurred (in brackets) and total days duration (‘130d’) Arcs have zero duration. Yellow (red) orange text indicates standard (strong) interactions. The topmost node shows the drugs prescribed at the start of the extract period; the double-bordered (green) node those at the end (reached by following the arcs).

to be part of an interactive tool to allow clinicians to control the useful level of generalisation.

We plan to develop more rigorous approaches to generalisation using metrics to measure the divergence in behaviour from that recorded in the event (prescription) log as nodes and arcs are removed or consolidated. This is the practice with other process representations (e.g. [21], [43]) but will be adapted to our probabilistic models and effect of grouping medication, and closely related to generalising probabilistic automata [23].

#### IV. POLYPHARMACY RESULTS

We outline here the initial findings drawn from the data analysis. An initially surprisingly high prevalence of polypharmacy is revealed. In total there were one or more interactions for almost 62% of the patients (approximately 54% with drugs with standard, 38% with strong interactions). These represent up to 50 and 16 unique combinations of standard- and strongly-interacting drugs respectively per patient. Overall there were 2,672 distinct pairs of drugs prescribed having a

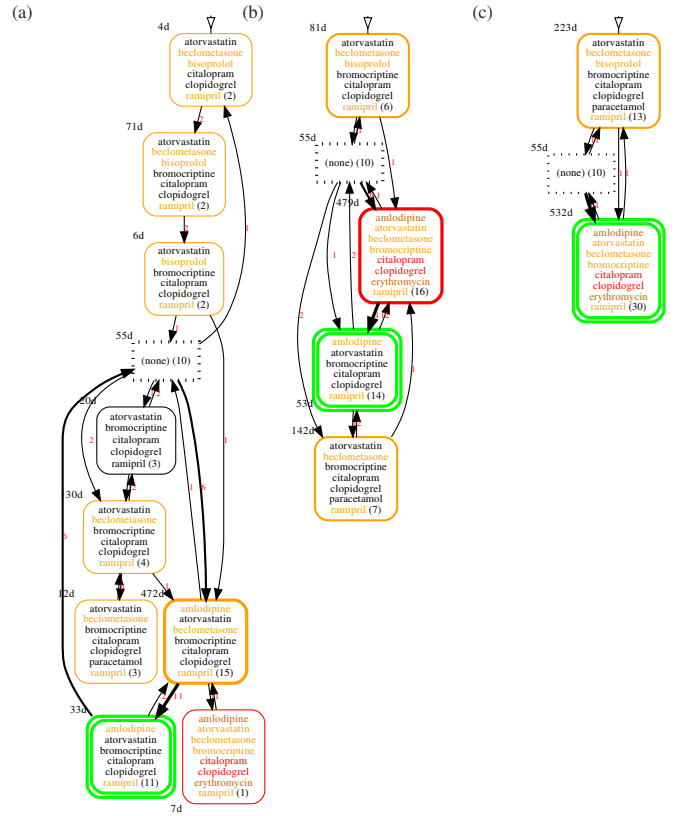


Fig. 3: Example of domain-specific generalisation, for the prescription process shown in Fig. 2(b): (a) after merging short-duration and small-group nodes into their neighbours; (b) after also merging neighbouring nodes representing similar groups of medications; and (c) after repeated merging with reducing threshold of similarity between groups.

standard interaction, 784 with a strong interaction. Although these figures seem very large, there are clinical reasons why some level of interaction between drugs may be necessary and accepted by clinicians – for instance, tolerance by the patient and mitigation of a more serious problem or interaction. Uncertainties such as interpreting the exact time when a prescription begins may also be responsible. Table V lists the top ten strongly-interacting drugs, by the number of patients prescribed and by the mean duration of prescription. The top pair, Simvastatin and Amlodipine, are treatments to reduce cholesterol and treat high blood pressure, respectively, and guidelines exist on dose limitations when taken together. This suggests that further information is needed to understand the true severity and impact of the interactions reported. We hypothesize that this may in some cases be obtainable from the prescription text, and also from expert knowledge.

Table IV reports the main summary statistics. The cleaned data contained 1,076 different drugs prescribed to 8,481 patients, with almost 95,000 different groups of concurrent medication prescribed. 639,638 prescription records were consolidated to about 250,000 by merging sequential and duplicate prescriptions (Section III-D). Individual patients were prescribed with between 1 and 45 different drugs, via up to



Per-Patient	Min.	Max.	Mean	Total
Prescription Events	1	1,568	75.42	639,638
Consolidated Prescriptions	1	268	29.54	250,513
Unique Medications	1	45	7.68	1,076
Unique Prescriptions	1	209	14.84	94,735
Standard Interactions	0	1,643	36.67	310,958
Strong Interactions	0	819	12.65	107,255
Unique Standard Interactions	0	50	2.39	2,672
Unique Strong Interactions	0	16	0.81	784

TABLE IV: Summary statistics for polypharmacy analysis. ‘Min.’, ‘Max.’ and ‘Mean’ are calculated per patient (from 8,481 patients), ‘Total’ across the whole patient cohort.

Drug	Patients Drug	Drug	Days Drug
amlodipine	355 simvastatin	alfuzosin	192.00 felodipine
furosemide	227 ramipril	alfuzosin	146.50 cardioplen xl
bendroflu- -methiazide	226 ramipril	adizem-xl	104.12 neбиволol
amlodipine	147 doxazosin	adizem-xl	88.67 metoprolol
indapamide	135 ramipril	citalopram	87.25 warfarin
citalopram	116 naproxen	eprosartan	82.60 indapamide
clarithromycin	111 simvastatin	methadone	80.75 olanzapine
amitriptyline	110 tramadol	eperenone	77.00 valsartan
amitriptyline	107 gabapentin	bendroflu- -methiazide	76.00 valsartan
indapamide	106 simvastatin	haloperidol	63.58 valproic acid

TABLE V: The ten most frequent strong interactions identified, by total number of patients prescribed (left) and mean duration (right).

1,568 (unconsolidated) prescriptions.

Fig. 4(a) shows that there is a ‘short tail’ of such very large numbers of prescriptions in the data set (occurring for each site). A similar profile is seen for the numbers of interactions. However although the patient with the largest number of prescriptions over the time period of the data extract also has the largest number of interactions, the Pearson’s coefficient of correlation of standard (strong) interactions with prescription count is only 0.52 (0.46). (Each data series in Fig. 4(a) is sorted independently.) In this case, the 1,568 prescriptions reflect up to 31 prescriptions issued on 157 days. Further investigation is needed to determine underlying reasons.

Figs. 4(b) and (c) confirm that there is some correlation between the number of prescriptions for a patient and the number of unique interactions (no duplicates), and between the numbers of standard and strong interactions for a patient. A fuller analysis of these results, including by site, demographics and patient condition, will be published elsewhere.

## V. DISCUSSION

The analysis showed the feasibility of applying a process mining approach to this type of data, despite the concept of a ‘prescription process’ being loosely defined and differing from the usual conception of a process, for example having a start and end point defined only by the parameters of the data extract. The resulting models exhibit a variety of process-like structures, and the initial polypharmacy results are of great interest to begin to answer the questions raised by the

protocol [5]. Detailed analysis will now be carried out, to relate polypharmacy to different patient cohorts, sites, clinical conditions, and to investigate root causes.

Our aim in using a probabilistic representation is that machine learning methods can be applied in a well-founded manner. We hope to apply clustering and clique detection to per-patient and multi-patient models, respectively, to discover common patterns of prescription that will be informative for clinicians in understanding the underlying reasons leading to polypharmacy. Our initial explorations in this direction have shown that such patterns are elusive, as hinted by the large numbers of different prescription and interaction combinations. We hope that again the involvement of clinician process and medication experts will enable appropriate methods to be developed and conclusions drawn.

Effective application of process mining and analysis of resultant models in the context of ‘noisy’ data and processes remains an open question. In fields with particularly heterogeneous data, flexible and variable processes such as healthcare, and more generally ‘knowledge work’ and Adaptive Case Management, naïve application of process mining algorithms results in ‘spaghetti’ (impossibly complex) models. We suggest that data-specific complexity-reduction and generalisation methods have a place in controlling complexity in a manner sympathetic to the domain in which process mining is applied. Our approach was to deal with concurrency implicitly for both mining and visualisation/generalisation. We plan to develop this as a more general method for dealing with loosely-defined (goal-oriented) processes.

Despite the useful results obtained from our method, and the promise of ‘big data’ [33] in healthcare [30], [26] more generally to enable evidence-based improvements to healthcare, we encountered considerable difficulties in gathering, cleaning, processing and interpreting data from multiple sources. Many of these difficulties seem symptomatic of a fragmented and evolving approach to data collection, presentation and storage, and could be significantly mitigated by relatively simple changes and adopting a mindset of curating data for the purposes of automated analysis. This can be seen in three main areas.

Firstly, text analytics allow information to be retrieved from free text, but a relatively large proportion (20% of records) could not be processed, potentially reducing the accuracy of the results. These issues included data errors, diminishing returns in modelling all nuances of the free text language used, and vague or missing information (such as the instruction to ‘take as directed’). Adopting a data and system design strategy of curating data for automated analysis would allow these exclusions to be significantly reduced and enable other useful analyses to be conducted more effectively in the future. Secondly, the availability of online information on drug interactions facilitated our analysis, but was hampered by frequent changes to the websites, partly due to evolving clinical information, but also again due to a lack of design for automated analysis. An adoption of the expectation that all data will be used (and useful) for automated analysis, stan-

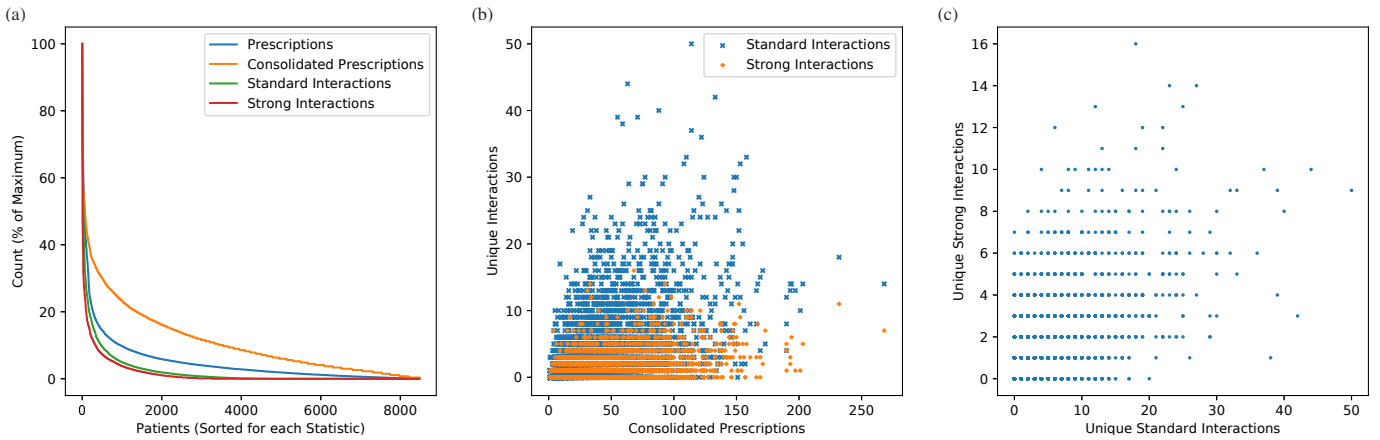


Fig. 4: Summary per-patient statistics on prescriptions: (a) scaled counts of prescriptions (before and after consolidation), standard and strong interactions, (b) counts of interactions against prescriptions, (c) some correlation (Pearson=0.63) between standard and strong interactions.

standardisation of formats and control of data revision processes, and consistency of naming conventions, would all improve the accuracy and efficiency of the analysis. Finally, clearly defined data semantics is also crucial. For example, at the process mining stage we assumed prescriptions to start on the ‘Date of Issue’, but this could be inaccurate, increasing the complexity of the mined models. Expert knowledge may mitigate this, but clear data descriptions should be provided to preserve data decisions and interpretations for future analyses.

## VI. LIMITATIONS AND FUTURE WORK

This study developed organically out of a need to answer particular questions from a given data set. As such it has several limitations which we aim to address in future work. Firstly, without a ground truth for the text analysis the evaluation of the grammar, although by experts, is necessarily subjective. The insights gained through this analysis would enable generation of hand-annotated or simulated data sets. Careful consideration will be necessary to ensure these are not overly-optimistic of the quality of data generated.

Secondly, a quality metric is required to measure and compare the generated process models, to enable well-founded generalisation. Since concurrency is implicit in the nodes, the present approaches to merging nodes to reduce visual complexity increase the concurrency, which may not be valid. Automata learning theory [10], [23] suggests starting points for such metrics [59]. At present, clinical experts are evaluating the medical conflicts retrieved, and will also guide the development of the model representation and a tool to support interactive visualisation of results. This assessment will be formalised and brought into the wider context of detecting and resolving conflicts between clinical pathways [60].

A limitation of the process mining techniques so far developed is the very large number of potential ‘activities’ (2,360 possible drugs, 1,076 seen in prescriptions). Even after aggregation into nodes representing groups of drugs, we still obtain ‘spaghetti’ models when mining models for multiple patients’ prescriptions. Application to a wider set of condi-

tions, and also to secondary care will add further complexity, not only in increased numbers of drugs and interactions, but also due to different drug formularies. Process mining is a visual technology, most usefully applied in close collaboration between (in our case) clinical and process mining experts. We plan to involve clinicians, pharmacists and other experts in developing the most useful representations and visualisations. There may also be further domain-specific information available from analysis of the GP comments, e.g. on shared decision-making with the patient, which can be used to interpret the relevance of discovered polypharmacy and hence potentially reduce complexity of the discovered processes.

Finally, additional data is needed to connect prescription processes and drug interactions discovered, with changing patient conditions. We hope to develop the analysis in this way using longitudinal data on patient disease coding.

## VII. CONCLUSION

Certain age groups and clusters of diseases place patients at increased risk. Little is known of the effect of polypharmacy and patterns of prescribing on multimorbid patients [49]. Our initial analysis showed that almost 62% of patients in our sample were prescribed drugs for which some form of interaction is known; strongly interacting for nearly 40%. Since polypharmacy and difficulties in de-prescribing are widely acknowledged [34], [58], [31], [14], this work and the further analysis it enables have potentially enormous benefit for healthcare providers, users and the wider healthcare economy.

We showed that process mining is a suitable approach to identify the underlying prescription processes leading to polypharmacy. Text and other data analytics tools enabled data to be brought together and interpreted from a number of sources, and shortcomings in data formats and documentation to be, to an extent, circumvented. The full analysis will facilitate a more tailored approach [24] relatable to groups of patients or combinations of morbidities, to help support the decision-making process of which drugs can be usefully de-prescribed – there is evidence that patients want to reduce their

medication [19], [45], however previous attempts to reduce the numbers of inappropriate prescriptions have failed [25].

We plan to integrate the analysis with our ongoing work in detecting and resolving conflicts between care pathways [60], to identify and use common patterns in prescription to understand the root causes leading to polypharmacy, and to recommend process changes to reduce its effect and improve the efficiency of prescribing and patient experience. We also plan to develop and apply the process mining approach for wider application to loosely-defined, goal-oriented processes.

#### ACKNOWLEDGMENT

This research was funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/M014401/1. The custom code for the data extract was funded by the University of Birmingham College of Medical and Dental Science Research Development Fund.

#### REFERENCES

- [1] *Proc. IEEE Symposium on Computational Intelligence and Data Mining (CIDM), part of the IEEE Symposium Series on Computational Intelligence (SSCI), Paris, France*. IEEE, 2011.
- [2] A. Adriansyah and J. C. A. M. Buijs. Mining process performance from event logs. In Rosa and Soffer [42], pp. 217–218.
- [3] A. Adriansyah, J. Munoz-Gama, J. Carmona, B. F. van Dongen, and W. M. P. van der Aalst. Alignment based precision checking. In Rosa and Soffer [42], pp. 137–149.
- [4] G. Alonso, P. Dadam, and M. Rosemann, eds. *Proc. Business Process Management (BPM), Brisbane, Australia*, LNCS 4714. Springer, 2007.
- [5] R. Backman, P. Weber, A. M Turner, M. Lee, and I. Litchfield. Assessing the extent of drug interactions among patients with multimorbidity in primary and secondary care in the West Midlands (UK): a study protocol for the Mixed Methods Multimorbidity Study (MiMMS). *BMJ Open*, 7(9):e016713 (8 pp.), 2017.
- [6] R. P. J. C. Bose and W. M. P. van der Aalst. Abstractions in process mining: A taxonomy of patterns. In Dayal et al. [12], pp. 159–175.
- [7] R. P. J. C. Bose and W. M. P. van der Aalst. Process mining applied to the BPI challenge 2012: Divide and conquer while discerning resources. In Rosa and Soffer [42], pp. 221–222.
- [8] M. Bozkaya, J. Gabriels, and J. M. E. M. van der Werf. Process diagnostics: A method based on process mining. In A. Kusiak and Sang-goo Lee, eds., *International Conference on Information, Process, and Knowledge Management (eKNOW), Cancun, Mexico*, pp. 22–27. IEEE Computer Society, 2009.
- [9] A. Burattin and A. Sperduti. Heuristics Miner for Time Intervals. In *Proc. ESANN*, Bruges, Belgium, 2010.
- [10] R. C. Carrasco and J. Oncina. Learning stochastic regular grammars by means of a state merging method. In R. C. Carrasco and J. Oncina, eds., *ICGI*, LNCS 862:139–152. Springer, 1994.
- [11] A. Dagliati, L. Sacchi, A. Zambelli, V. Tibollo, L. Pavesi, J. H. Holmes, and R. Bellazzi. Temporal electronic phenotyping by mining careflows of breast cancer patients. *Journal of Biomedical Informatics*, 66:136–147, 2017.
- [12] U. Dayal, J. Eder, J. Koehler, and H. A. Reijers, eds. *Proc. Business Process Management (BPM), Ulm, Germany*, LNCS 5701. Springer, 2009.
- [13] M. de Leoni, W. M. P. van der Aalst, and B. F. van Dongen. Data- and resource-aware conformance checking of business processes. In W. Abramowicz, D. Kriksciuniene, and V. Sakalauskas, eds., *Proc. BIS, LNBIP 117:48–59*. Springer, 2012.
- [14] S. Dumbreck, A. Flynn, M. Nairn, M. Wilson, S. Treweek, S. W Mercer, P. Alderson, A. Thompson, K. Payne, and B. Guthrie. Drug-disease and drug-drug interactions: systematic examination of recommendations in 12 UK national clinical guidelines. *BMJ* 350:h949 (8 pp.), 2015.
- [15] C. Fernández-Llatas, A. Martínez-Millana, A. Martínez-Romero, J. Benedí, and V. Traver. Diabetes care related process modelling using process mining techniques. lessons learned in the application of interactive pattern recognition: coping with the spaghetti effect. In *Proc. International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, Milan, Italy*, pp. 2127–2130. IEEE, 2015.
- [16] D. R. Ferreira and D. Gillblad. Discovering process models from unlabelled event logs. In Dayal et al. [12], pp. 143–158.
- [17] D. R. Ferreira, M. Zacarias, M. Malheiros, and P. Ferreira. Approaching process mining with sequence clustering: Experiments and findings. In A. et al. [4], pp. 360–374.
- [18] The King’s Fund. Polypharmacy and medicines optimisation. Online, 2013. Martin Duerden and Tony Avery and Rupert Payne.
- [19] A. Galazzi, M. Lusignani, M. T. Chiarelli, P. M. Mannucci, C. Franchi, M. Tettamanti, E. Reeve, and A. Nobili. Attitudes towards polypharmacy and medication withdrawal among older inpatients in Italy. *International Journal of Clinical Pharmacy*, 38(2):454–461, 2016.
- [20] G. Greco, A. Guzzo, and L. Pontieri. Discovering expressive process models by clustering log traces. *IEEE Trans. Knowl. Data Eng.*, 18(8):1010–1027, 2006.
- [21] C. W. Günther and W. M. P. van der Aalst. Fuzzy mining – adaptive process simplification based on multi-perspective metrics. In Alonso et al. [4], pp. 328–343.
- [22] B. Guthrie, C. McCowan, P. Davey, C. R Simpson, T. Dreischulte, and K. Barnett. High risk prescribing in primary care patients particularly vulnerable to adverse drug events: cross sectional population database analysis in Scottish general practice. *BMJ*, 342:d3514, 2011.
- [23] S. Jacquemont, F. Jacquenet, and M. Sebban. Mining probabilistic automata: A statistical view of sequential pattern mining. *Machine Learning*, 75(1):91–127, 2009.
- [24] C. Jäger, J. Steinhäuser, T. Freund, S. Kuse, J. Szecsenyi, and M. Wensing. A tailored programme to implement recommendations for multimorbid patients with polypharmacy in primary care practices – process evaluation of a cluster randomized trial. *Implementation Science*, 12(1):31, Mar 2017.
- [25] A. Kolhatkar, L. Cheng, F. K. I. Chan, M. Harrison, and M. R. Law. The impact of medication reviews by community pharmacists. *Journal of the American Pharmacists Association*, 56(5):513–520.e1, 2016.
- [26] S. C. Kruse, R. Goswamy, Y. Raval, and S. Marawi. Challenges and opportunities of big data in health care: A systematic review. *JMIR Med Inform*, 4(4):e38, 2016.
- [27] G. T. Lakshmanan, S. Rozsnyai, and F. Wang. Investigating clinical care pathways correlated with outcomes. In F. Daniel, J. Wang, and B. Weber, eds., *Proc. Business Process Management (BPM), Beijing, China*, LNCS 8094:323–338. Springer, 2013.
- [28] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst. Discovering block-structured process models from event logs - A constructive approach. In J. M. Colom and J. Desel, eds., *Proc. Application and Theory of Petri Nets and Concurrency (PETRI NETS), Milan, Italy, 2013*, LNCS 7927:311–329. Springer, 2013.
- [29] 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. *Comp. in Bio. and Med.*, 77:125–134, 2016.
- [30] J. Luo, Min Wu, D. Gopukumar, and Yiqing Zhao. Big data application in biomedical research and health care: A literature review. *Biomedical Informatics Insights*, 8:1–10, 2015.
- [31] R. L. Maher, J. Hanlon, and E. R. Hajjar. Clinical consequences of polypharmacy in elderly. *Expert Opinion on Drug Safety*, 13(1):57–65, 2014. PMID: 24073682.
- [32] R. Mans, W. M. P. van der Aalst, and R. J. B. Vanwersch. *Process Mining in Healthcare – Evaluating and Exploiting Operational Healthcare Processes*. Springer Briefs in BPM. Springer, 2015.
- [33] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers. *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute, 2011.
- [34] Z. A. Marcum and W. F. Gellad. Medication adherence to multidrug regimens. *Clinics in Geriatric Medicine*, 28(2):287–300, 2012.
- [35] S. Montani, M. Striani, S. Quaglini, A. Cavallini, and G. Leonardi. Knowledge-based trace abstraction for semantic process mining. In A. ten Teije, C. Popow, J. H. Holmes, and L. Sacchi, eds., *Proc. Artificial Intelligence in Medicine (AIME), Vienna, Austria*, LNCS 10259:267–271. Springer, 2017.

- [36] A. Nobili, G. Licata, F. Salerno, L. Pasina, M. Tettamanti, C. Franchi, L. De Vittorio, A. Marengoni, S. Corrao, A. Iorio, M. Marcucci, and P. M. Mannucci. Polypharmacy, length of hospital stay, and in-hospital mortality among elderly patients in internal medicine wards: the Reposit study. *European Journal of Clinical Pharmacology*, 67(5):507–519, 2011.
- [37] OMG. Business Process Model and Notation (BPMN), 2009.
- [38] A. Partington, M. T. Wynn, S. Suriadi, Chun Ouyang, and J. Karnon. Process mining for clinical processes: A comparative analysis of four Australian hospitals. *ACM Trans. Management Inf. Syst.*, 5(4):19:1–19:18, 2015.
- [39] M. Peleg, P. Soffer, and J. Ghattas. Mining process execution and outcomes – position paper. In A. H. M. ter Hofstede, B. Benatallah, and Hye-Young Paik, eds., *Proc. Business Process Management Workshops, BPM, Revised Selected Papers*, LNCS 4928:395–400. Springer, 2007.
- [40] A. Rebuge and D. R. Ferreira. Business process analysis in healthcare environments: A methodology based on process mining. *Inf. Syst.*, 37(2):99–116, 2012.
- [41] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro. Process mining in healthcare: A literature review. *Journal of Biomedical Informatics*, 61:224–236, 2016.
- [42] M. La Rosa and P. Soffer, eds.. *BPM Workshops – BPM 2012, Tallinn, Estonia, Revised Papers*, LNBIP 132. Springer, 2013.
- [43] A. Rozinat and W. M. P. van der Aalst. Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1):64–95, 2008.
- [44] J. Sinnige, J. C. Braspenning, F. G. Schellevis, K. Hek, I. Stirbu, G. P. Westert, and J. C. Korevaar. Inter-practice variation in polypharmacy prevalence amongst older patients in primary care. *Pharmacoepidemiology and Drug Safety*, 25(9):1033–1041, 2016. PDS-15-0391.R2.
- [45] C. Sirois, N. Ouellet, and E. Reeve. Community-dwelling older people’s attitudes towards deprescribing in Canada. *Research in Social and Administrative Pharmacy*, 13(4):864–870, 2017.
- [46] S. M. Smith, H. Soubhi, M. Fortin, C. Hudon, and T. O’Dowd. Managing patients with multimorbidity: systematic review of interventions in primary care and community settings. *BMJ*, 345:e5205, 2012.
- [47] N. Tax, N. Sidorova, R. Haakma, and W. M. P van der Aalst. Mining local process models. *J. Innovation in Digital Ecosystems*, 3:183–196, 2016.
- [48] P. Taylor, M. Leida, and B. Majeed. Case study in process mining in a multinational enterprise. In K. Aberer, E. Damiani, and T. Dillon, eds., *Data-Driven Process Discovery and Analysis*, LNBIP 116:134–153. Springer Berlin Heidelberg, 2012.
- [49] C. Tsang, A. Bottle, A. Majeed, and P. Aylin. Adverse events recorded in English primary care: observational study using the general practice research database. *British Journal of General Practice*, 63(613):e534–e542, 2013.
- [50] W. M. P. van der Aalst. Process mining: Discovering and improving spaghetti and lasagna processes. In *CIDM* [1], pp. 13–20.
- [51] W. M. P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, Heidelberg Dordrecht London New York, 2011.
- [52] W. M. P. van der Aalst, A. Adriansyah, and B. van Dongen. Replaying history on process models for conformance checking and performance analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(2):182–192, 2012.
- [53] W. M. P. van der Aalst, T. Weijters, and L. Maruster. Workflow mining: Discovering process models from event logs. *IEEE Trans. Knowl. Data Eng.*, 16(9):1128–1142, 2004.
- [54] W. M. P. van der Aalst. The Application of Petri Nets to Workflow Management. *Journal of Circuits, Systems, and Computers*, 8(1):21–66, 1998.
- [55] W. M. P. van der Aalst, M. H. Schonenberg, and Minseok Song. Time prediction based on process mining. *Inf. Syst.*, 36(2):450–475, 2011.
- [56] Takashi Washio and Hiroshi Motoda. State of the art of graph-based data mining. *SIGKDD Explorations*, 5(1):59–68, 2003.
- [57] M. Wauters, M. Elseviers, B. Vaes, J. Degryse, O. Dalleur, R. Vander Stichele, L. Van Bortel, and M. Azermi. Polypharmacy in a Belgian cohort of community-dwelling oldest old (80+). *Acta Clinica Belgica*, 71(3):158–166, 2016. PMID: 27105401.
- [58] M. Wauters, M. Elseviers, B. Vaes, J. Degryse, O. Dalleur, R. Vander Stichele, T. Christiaens, and M. Azermi. Too many, too few, or too unsafe? impact of inappropriate prescribing on mortality, and hospitalization in a cohort of community-dwelling oldest old. *British Journal of Clinical Pharmacology*, 82(5):1382–1392, 2016. MP-00234-16.R1.
- [59] P. Weber, B. Bordbar, and P. Tiño. A framework for the analysis of process mining algorithms. *IEEE T. Systems, Man, and Cybernetics: Systems*, 43(2):303–317, 2013.
- [60] P. Weber, J. B. F. Filho, B. Bordbar, M. Lee, I. Litchfield, and R. Backman. Automated Conflict Detection Between Medical Care Pathways. *Journal of Software: Evolution and Process (Special Issue on Software Engineering for Connected Health)*, 2017.
- [61] A. J. M. M. Weijters and J. T. S. Ribeiro. Flexible Heuristics Miner (FHM). In *CIDM* [1], pp. 310–317.