Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Special Communication

Process mining-driven analysis of COVID-19’s impact on vaccination patterns

Adriano Augusto, Timothy Deitz, Noel Faux, Jo-Anne Manski-Nankervis, Daniel Capurro

The University of Melbourne, Australia

ARTICLE INFO

Keywords:
Process mining
Data analysis
Healthcare processes
Vaccination
COVID-19

ABSTRACT

Process mining is a discipline sitting between data mining and process science, whose goal is to provide theoretical methods and software tools to analyse process execution data, known as event logs. Although process mining was originally conceived to facilitate business process management activities, research studies have shown the benefit of leveraging process mining in healthcare contexts. However, applying process mining tools to analyse healthcare process execution data is not straightforward. In this paper, we show a methodology to: i) prepare general practice healthcare process data for conducting a process mining analysis; ii) select and apply suitable process mining solutions for successfully executing the analysis; and iii) extract valuable insights from the obtained results, alongside leads for traditional data mining analysis. By doing so, we identified two major challenges when using process mining solutions for analysing healthcare process data, and highlighted benefits and limitations of the state-of-the-art process mining techniques when dealing with highly variable processes and large data-sets. While we provide solutions to the identified challenges, the overarching goal of this study was to detect differences between the patients’ health services utilization pattern observed in 2020—during the COVID-19 pandemic and mandatory lock-downs—and the one observed in the prior four years, 2016 to 2019. By using a combination of process mining techniques and traditional data mining, we were able to demonstrate that vaccinations in Victoria did not drop drastically—as other interactions did. On the contrary, we observed a surge of influenza and pneumococcus vaccinations in 2020, as opposed to other research findings of similar studies conducted in different geographical areas.

1. Introduction

The discipline of Process Mining [1] was born with the goal to design automated data analysis techniques that could support the phases of the business process management lifecycle [2], especially those phases where the data analysis plays a central role, e.g., process discovery and process monitoring. Over the past two decades, research in the area of process mining has generated a number of methodologies and software tools (henceforth, process mining techniques) [3–6]. Process mining techniques usually require process execution data, which is known as event logs. It is possible to distinguish two major families of process mining techniques [2]: operational techniques and tactical techniques. The former family encompasses techniques whose goal is to generate insights in real-time during the process execution, e.g., estimating the probability of a negative event to happen; or the likelihood of a specific process outcome. The latter family encompasses techniques whose goal is to help analysts to discover, analyse, and periodically monitor the process execution in order to understand how the process is performed, what are its weaknesses, and how the process can be improved. Two of the most popular tactical process mining techniques, which we will refer throughout this study, are: i) automated process discovery – which allows to automatically discover a process model from event logs; ii) variant analysis – which facilitates the analysis of behavioural differences between process variants (e.g., process instances with a positive outcome versus those with a negative outcome).

Although process mining was initially conceived to be applied within business contexts (such as, banking, wholesaling, manufacturing), research has shown that its value can be harnessed and reused in a multitude of different contexts, including healthcare [7–12]. This study sits within the healthcare context, and it is set during the COVID-19 pandemic in Victoria (Australia). Our research goal was to identify differences between the patients’ health services utilization pattern observed in 2020—during the COVID-19 pandemic and mandatory lock-downs—and the one observed in the prior four years, 2016 to 2019.
Given that health services are provided via the enacting of healthcare processes, process mining techniques are ideal for achieving our research goal, in particular, process discovery and process variant analysis techniques. To this end, we analysed process execution data extracted from more than 100 general practice (GP) clinics in Victoria. This data included more than 30 million events capturing the GP healthcare processes of more than one million patients in Victoria, over a time-span of approximately five years. While, in this study, we do not build on top of our findings to lead process improvement initiatives, given that no clinics or hospitals were directly involved, it is worth noting that the outcome of our process analysis can be exploited to improve healthcare services during emergency periods, for instance, enhancing healthcare resources allocation (including personnel, medications, vaccines, and services) and ensuring short processing times.

The main contributions of this study are the following.

- The showcasing of a methodology to: i) prepare general practice healthcare process data for conducting a process mining analysis; ii) select and apply suitable process mining solutions for successfully executing the analysis; and iii) extract valuable insights from the results of the process mining analysis alongside leads for traditional data mining analysis.
- The identification of two major challenges when dealing with general practice healthcare process data: improper timestamp granularity; and unbounded process traces, which have no explicit start and end.
- A novel method for fixing timestamp equivalence issues in process execution data.
- A novel method to identify boundaries of incomplete traces with unknown start events.
- A critical assessment of the capabilities of state-of-the-art process mining techniques as well as their limitations when dealing with large data-sets recording highly variable processes – which is typical of the healthcare processes. This allowed us to draw directions for future research in the area of applied process mining in healthcare.
- From a medical perspective, the results of our analysis show that vaccinations in Victoria did not drop as drastically as other clinical interactions did. On the contrary, we observed a surge of influenza and pneumococcus vaccinations in 2020. Also, these findings differ from other research findings of similar studies conducted in different geographical areas in the equivalent seasonal periods [13–15], providing a different perspective.

The remainder of the paper is structured as follows. In Section 2 we discuss related work and background. In Section 3, we describe the data, the analysis we ran, the challenges we faced, and the solution we adopted. In Section 4, we review the findings of the data analysis, providing a medical interpretation and considering their consequences. Lastly, Section 5 summarises our results and draws the conclusion.

2. Background and related work

2.1. Process mining in healthcare

The analysis of healthcare delivery from the process perspective has been a core aspect of health services research and redesign. However, until recently, the analysis of healthcare data using a process perspective has been challenging due to the limited availability of electronic health data (and/or its poor quality) and the lack of powerful methods to quickly make sense of it [16,12]. The recent adoption of electronic health records alongside process-aware information systems [17] has generated vast amounts of healthcare data—both clinical and administrative—that can be leveraged to better understand healthcare processes. Several systematic reviews have highlighted the use and potential benefits of applying process mining methods to understand and improve healthcare processes [7,18,19,12], and research reports include uses of a range of process mining techniques such as automated process discovery, conformance checking, and process variant analysis.

Automated process discovery techniques [20–23,3] allow one to discover patients’ clinical pathways from the recordings of their healthcare process activities captured by the hospitals and clinics information systems [24,16]. Conformance checking techniques [25,5] allow one to automatically compare the observed healthcare process behaviour (in the form of process execution data) against a prescribed process behaviour to identify differences between actual and normative healthcare behaviour. The latter is usually provided in the form of an imperative process model or as a set of declarative process rules [26], which rather than capturing the full process behaviour may describe clinical guidelines. Process variant analysis techniques [27,28,4,29] allow one to automatically compare two or more sets of healthcare process executions exhibiting different outcomes (or performance) to identify relevant differences between the executions that may have had an impact on the outcome or performance of the healthcare process. These type of techniques are applied to answer questions such as: what were the differences between the healthcare treatments provided by two different hospitals to patients having the same diagnosis?

One of the earliest application of process mining in healthcare dates back to 2008, Mans et al. [24] used Heuristics Miner [20] to extract insights from healthcare process data, both from clinical and administrative perspective, including process handovers analysis by leveraging the process mining analytics platform ProM. Poelmans et al. [30] used a combination of process mining and data mining techniques to detect and analyse differences in the healthcare pathways of patients treated for breast cancer and how they would respond to different therapies. Lakshmanan et al. [31] proposed an approach for discovering patients healthcare pathways and correlate them to their outcomes, combining techniques from process mining and data mining (including clustering and pattern mining). Suriadi et al. [32] applied process mining techniques to understand the differences of the treatments provided to patients suffering from chest pain at four South Australian hospitals. Partington et al. [33] applied process mining techniques to analyse the quality and the costs of the healthcare services provided to patients at one South Australian hospital. Roviani et al. [26] reported a case study on how to leverage declarative process mining techniques to identify divergences between clinical guidelines and the observed execution of clinical processes, at the urology department of the Isala hospital in the Netherlands. Leonardi et al. [8] proposed a method to abstract low-level process execution data (in the form of simple actions), turning it into high-level data that can be used for process mining applications. They validated their method by discovering process models from healthcare services, showing that their method improved the graphical representation of the healthcare processes, and facilitated the clustering of similar process executions. Alvarez et al. [9] applied process mining techniques to discover process models capturing how healthcare professionals operate within emergency rooms, analysing them to identify opportunities for process improvement. Chen et al. [10] proposed a framework to extract high-level descriptions of medical treatment processes from electronic medical records by applying clustering techniques on doctor order set sequences. Their framework allows to enrich the extracted process descriptions with additional information regarding the process performance (e.g., cost, length), providing support for improvement. Yang et al. [11] designed a process mining approach to automatically and in real-time detect process deviations from recommended clinical guidelines. They validated their approach on a set of pediatric trauma resuscitation procedures, demonstrating the effectiveness of their solution.

All these studies on process mining in healthcare represent only a fraction of the existing ones, but reporting on all of them would require a separate study and it would be outside the scope of this one. Hence, we

1 https://www.promtools.org/.
refer the interested reader to the latest literature reviews [19,18].

Given the diversity of tools available and the applicability of process mining to healthcare, we used this perspective to understand changes in health services utilization patterns during the COVID-19 pandemic in Australia.

### 2.2. Process changes during the COVID-19 pandemic

Since the early months of the COVID-19 pandemic, the main drivers behind lock-downs and stay-at-home measures were the need to reduce face-to-face interactions to prevent the virus from spreading uncontrollably, the subsequent increase in morbidity, mortality, and overwhelming of healthcare service providers.

In parallel, there were growing concerns that stay-at-home recommendations, lockdown measures, and the fear of becoming infected would have a deep impact on the provision of non-COVID-19 health services. Although heterogeneous, most governments across the world recommended some form of mobility reduction measures to reduce the transmission rate of SARS-CoV-19 so the expectations were that most countries would be impacted, although at different extents. Several publications reported the observed effects on the utilization of health services. The World Stroke Organization reported on a reduction in the number of patients being diagnosed with stroke despite COVID-19 apparently increasing the risk of this diseases and attributed the change to reduced access to health services [34]. These findings were confirmed in the USA [35]. Similar effects were described for patients with acute myocardial infarction [36], and cancer [37,38], among other conditions. This phenomenon was also observed for preventative care services such as cancer screening [39-41], and maternal and child health services [42]. In particular, there were growing concerns that a significant reduction in immunizations would result in an increase in vaccine-preventable conditions [43,13].

The goal of this study was to analyse changes in health services utilization patterns during the 2020 COVID-19 pandemic and associated lock-downs in Victoria (Australia).

### 3. Analysis, observations, and challenges

In this section, we introduce the data we analysed, discussing its characteristics and highlighting those that are the most critical in the context of this study. We describe what methodology and tools we used to analyse the data, what findings we uncovered, what challenges we faced during the analysis and how we addressed them. While we were able to solve some of these challenges, by proposing approaches that can be reused in different contexts, other challenges remain open or partially addressed and should be considered in future research work in the area of process mining.

#### 3.1. Preliminaries

Before discussing our analysis, we provide some formal definition for the concepts we refer to throughout section. While we contextualised these definitions within our study, we remark that these are well-known definitions and concepts in the area of process mining [1].

**Definition 1. Event** – An event \( e \) captures the execution of an activity within a process instance. An event can be represented as a tuple \((x_1, x_2, \ldots, x_n)\), where each element \( x_i \) captures an attribute of the event, and at least three attributes are present: the process instance ID (\( c – \) event ID); the label of the activity the event refers to (\( a – \) event activity); and the timestamp (\( t – \) event timestamp). Additional attributes usually capture the process resource who executed the activity, customer information, etc. In the following, given an event \( e \), we will refer to its three required attributes with the notation \( e_c, e_a, e_t \).

**Definition 2. Event Log** – An event log \( \mathcal{L} \) is a sequence of events \((e_1, e_2, \ldots, e_n)\), such that all the events are ordered by their timestamp. Formally, \( \forall e_i \in \mathcal{L}[i \in [1, n-1] \land N(\mathit{e}_t^{i+1} \notin \mathit{e}_t^{i})] \land N(\mathit{e}_c^{i+1} \notin \mathit{e}_c^{i}) \land N(\mathit{e}_a^{i+1} \notin \mathit{e}_a^{i}) \).

**Definition 3. Trace** – Given an event log \( \mathcal{L} \), a trace of the event log \( t \in \mathcal{L} \) is a sequence of events, \( t = (e_1, e_2, \ldots, e_n) \) such that all the events belong to the event log, all the events are ordered by their timestamp, and all the events have the same event ID attribute. Formally, \( \forall e_t \in t[i \in [1, n-1] \land N(\mathit{e}_t^{i+1} = \mathit{e}_t^{i})] \land N(\mathit{e}_c^{i+1} = \mathit{e}_c^{i}) \land N(\mathit{e}_a^{i+1} = \mathit{e}_a^{i}) \).

We note that, according to Definition 3, we can also consider an event log as a multiset of traces.

**Definition 4. Directly-follows Relation** – Given an event log \( \mathcal{L} = (e_1, e_2, \ldots, e_n) \), we say that a directly-follows relation holds between any two events \( e_i, e_j \) if and only if \( e_i \) and \( e_j \) belong to the same trace and \( j = i + 1 \), in other words, the two events follow each other in (at least) one trace. We indicate such a relation with the notation \( e_i \rightarrow e_j \). Formally, given \( e_i, e_j \in \mathcal{L} \), \( e_i \rightarrow e_j \Leftrightarrow \exists e_i \in \mathcal{L}[i \leq j \land e_j \in \mathcal{L}[j+1] \land j = i + 1 \). We extend the concept of directly-follows relation to the event activities, i.e., if \( e_i \rightarrow e_j \) then we say that also \( e_a_i \rightarrow e_a_j \) holds.

**Definition 5. Directly-Follows Graph (DFG)** – Given an event log \( \mathcal{L} \), its Directly-Follows Graph (DFG) is a directed graph \( G = (N,E) \), where: \( N \) is the set of nodes, \( N = \{ n \mid \exists e \in \mathcal{L} \land e_a = x \} \); and \( E \) is the set of edges \( E = \{ (x,y) \in N \times N | \exists e_1, e_2 \in \mathcal{L} \land e_1 \rightarrow e_2 \land e_1a = x \land e_2a = y \} \). In other words, each node of the DFG represents a unique activity recorded in the event log, and each edge of the DFG represents a directly-follows relation between two activities – represented by the source node and target node of the edge.

**Definition 6. (Business) Process** – A (Business) Process is a sequence of events, activities, and decisions involving actors and data objects triggered by a specific start event and leading to a specific end event (i.e., process outcome) that delivers value to a customer.

#### 3.2. Dataset

In this study we used the Patron dataset [44]. This dataset stores de-identified patient data from the Patron primary care data repository (extracted from consenting general practices), that has been created and is operated by the Department of General Practice at The University of Melbourne [44,45]. This dataset is aggregated from more than 100 General Practice (GP) clinics in Victoria (Australia) and includes both administrative and clinical data, including all interactions between patients and their GPs, for more than one million patients. Access to the data was approved by the Melbourne Health Human Research Ethics Committee (HREC). The dataset is stored in a relational database, which includes the following six tables: Patient Details (Demographics); Patient Clinical Information; Medical History (Diagnoses); Patient Visits; Medications; Investigations (Pathology and Imaging). While the first three tables contain information regarding the patient and their clinical history; the last three tables contain information regarding the patient healthcare processes, respectively: information on patient visits to and interactions with their GP doctor(s); information on patient drugs prescriptions; and information on patient pathology and imaging tests and

---

**Table 1**

| Activity Label | Activity Description |
|---------------|----------------------|
| A             | Patient attends a GP doctor visit |
| B             | GP records a measurement (e.g., blood pressure) |
| C             | Patient is prescribed a medication |
| D             | Patient is prescribed a medication refill |
| E             | Patient is referred for a laboratory or imaging study (e.g., blood analysis) |
| F             | Tests results are recorded |
| G             | One or more vaccinations are administered/recorded |
results. Looking at the latter three tables through the lens of Definition 1, an event ID corresponds to a patient ID, which identifies a unique patient accessing GP services across all the tables. An event activity corresponds to a medical activity the patient underwent. From the three tables capturing the patient healthcare process, it is possible to extract seven medical activities, which are reported in Table 1. These activities capture ambulatory patients’ interactions with their GPs, including the drugs they have been prescribed, their pathology and imaging tests and results, and their vaccinations. For simplicity, we will refer to each of these seven activities by using a letter A to G (following the mapping in Table 1). Lastly, the event timestamp corresponds to the time a medical activity was completed. We note that the Patron dataset does not record information regarding activities’ lifecycle, e.g., the start and the completion of the activities, and that the timestamp granularity is at day-level (i.e., the smallest difference between timestamps is at day-level). Such a timestamp granularity is frequent in the healthcare contexts, and (at least in our case) it is related to how the system records events into the database. Consequently, it was virtually impossible to infer a better timestamp granularity (e.g., hours and minutes) or the duration of a single medical activity (e.g., how long a GP visit would last). In light of this, in the Patron dataset, a trace captures a unique patient accessing GP services over the time, i.e., a process instance of the GP day-to-day healthcare process.

Once the scope of our analysis was set, we moved to data extraction and processing. The Patron dataset, as mentioned above, already included all the data we required to analyse the selected process. The extraction of this data was performed outside this study, and it is not our contribution. However, healthcare process data rarely comes in the form provided in 2020 during the COVID-19 pandemic and mandatory lockdowns –and those observed in the prior four years, 2016 to 2019. The authors of this paper composed the research team, bringing expertise in process mining and analysis; evaluation; and process improvement and support. We thoroughly executed all the stages with the exception of the last stage. Given that this study did not involve GP clinics and healthcare practitioners, we did not have the means to implement a redesigned process, besides, it would have been outside the scope of this study.

### 3.4. Planning, data extraction and processing

Following the PM² methodology, we started from the planning, which includes three steps: i) selecting the process to analyse; ii) determining the process analysis goal; iii) and assembling a team. Indirectly, the Patron dataset drove the process selection. Given that it captures ambulatory patients’ interactions with their GPs, we selected for our analysis the GP day-to-day healthcare process. Our analysis objective was to identify differences between the GP healthcare services provided in 2020–during the COVID-19 pandemic and mandatory lockdowns –and those observed in the prior four years, 2016 to 2019. The authors of this paper composed the research team, bringing expertise in process mining, data mining, and (medical) general practice.

Once the scope of our analysis was set, we moved to data extraction and processing. The Patron dataset, as mentioned above, already included all the data we required to analyse the selected process. The extraction of this data was performed outside this study, and it is not our contribution. However, healthcare process data rarely comes in the form of ready-to-use event logs [16,33], which is the required data format for conducting a process mining analysis [1,46], and the Patron dataset was no exception. During this stage, we focused on transforming the available data into an event log that could allow us to achieve our analysis goal. This required us to identify what entries of the relational database were suitable to be turned into events. As mentioned above, we extract all the entries from three tables out of six, which captured the medical activities shown in Table 1. Each entry of a table included the patient ID and the timestamp, hence, the conceptual mapping from table entries to events was straightforward. We note that this mapping was facilitated by the existing of a very extensive data dictionary describing the Patron

### Table 2

| Event | Traces | Events | Trace length |
|-------|--------|--------|--------------|
| Log   | Total  | Distinct (#) (%) | Total  | Distinct | Filtered (#) (%) | Min | Avg | Max |
| GP16  | 507,075| 202,621 | 40.0    | 6,231,914| 7            | 232,586 | 3.7   | 1   | 11  | 652   |
| GP17  | 520,502| 215,079 | 41.3    | 6,696,341| 7            | 221,548 | 3.3   | 1   | 12  | 834   |
| GP18  | 531,618| 221,732 | 41.7    | 6,865,778| 7            | 229,803 | 3.3   | 1   | 12  | 748   |
| GP19  | 522,022| 221,789 | 42.5    | 6,886,762| 7            | 266,904 | 3.9   | 1   | 12  | 1966  |
| GP20  | 401,370| 167,107 | 41.6    | 5,106,686| 7            | 205,213 | 4.0   | 1   | 12  | 2317  |
| GP16–20 | 2,482,587| 1,028,328 | 41.4    | 31,769,481| 7            | 1,156,054 | 3.6   | 1   | 12  | 2317  |

### Table 3

Distribution of distinct traces over a set of frequency buckets.

| Freq. Bucket | 2020 (GP20 Log) | 2019 (GP19 Log) | 2018 (GP18 Log) | 2017 (GP17 Log) | 2016 (GP16 Log) |
|--------------|----------------|----------------|----------------|----------------|----------------|
| 1-5          | 164,113 | 98.2%   | 217,973 | 98.3%   | 217,829 | 98.2%   | 211,195 | 98.2%   | 198,769 | 98.1%   |
| 6-10         | 1,564   | 0.94%   | 1,986   | 0.90%   | 2,040   | 0.92%   | 2,017   | 0.94%   | 1,995   | 0.98%   |
| 11-15        | 442     | 0.26%   | 600     | 0.27%   | 649     | 0.29%   | 618     | 0.29%   | 630     | 0.31%   |
| 16-20        | 240     | 0.14%   | 269     | 0.12%   | 256     | 0.12%   | 307     | 0.14%   | 284     | 0.14%   |
| 21-50        | 437     | 0.26%   | 578     | 0.26%   | 557     | 0.25%   | 551     | 0.26%   | 543     | 0.27%   |
| 51-100       | 136     | 0.08%   | 161     | 0.07%   | 186     | 0.08%   | 177     | 0.08%   | 194     | 0.10%   |
| 101-500      | 130     | 0.08%   | 165     | 0.07%   | 160     | 0.07%   | 163     | 0.08%   | 154     | 0.08%   |
| 501-1000     | 21      | 0.03%   | 30      | 0.03%   | 26      | 0.02%   | 23      | 0.02%   | 24      | 0.03%   |
| 1,001-2,000  | 12      | 0.01%   | 10      | 0.01%   | 14      | 0.01%   | 13      | 0.01%   | 14      | 0.01%   |
| 2,001-5,000  | 5       | 0.01%   | 10      | 0.01%   | 8       | 0.01%   | 9       | 0.01%   | 8       | 0.01%   |
| 5,001-10,000 | 3       | 0.01%   | 2       | 0.01%   | 3       | 0.01%   | 2       | 0.01%   | 2       | 0.01%   |
| 10,001-20,000| 2       | 0.01%   | 4       | 0.01%   | 2       | 0.01%   | 2       | 0.01%   | 2       | 0.01%   |
| 20,001+      | 1       | 0.01%   | 1       | 0.01%   | 2       | 0.01%   | 2       | 0.01%   | 2       | 0.01%   |
The data extracted captured a time-span of (almost) five years, from January 2016 to November 2020, but we reduced this time-span to keep only the data collected between 01-March to 30-November for the years 2016, 2017, 2018, 2019, and 2020. This choice was driven by three factors: i) our analysis goal (as mentioned above); ii) a key date in the international response to the COVID-19 pandemic; iii) and our latest access to data. Precisely, given that the World Health Organization (WHO) officially declared the COVID-19 a pandemic on the 11th March, we set the start date of our analysis on the 1st of March, while for the end date we were forced to set it to the 30th of November, which was our latest available access to data. We also note that the two dates are closely related to the enforcement of the first lockdown restrictions in Victoria (16-March-2020) and the lifting of the last lockdown restrictions in Victoria (09-November-2020), in the year 2020.

The data was extracted via an ad-hoc R-script and saved in the form of CSV event logs. These CSV logs were then converted in the standard

---

**Fig. 1.** Activity frequencies, graphical comparison across years 2020–2016. \( A = \) GP visit, \( B = \) GP records measurement, \( C = \) medication prescribed, \( D = \) medication refill, \( E = \) lab/imaging referral, \( F = \) test results recorded, \( G = \) vaccination administered/recorded.
XES format via Apromore (academic version), which can be used without internet access and does not have limits on the amount of data to be processed, as opposed to Disco or Celonis. Alternatively, we could have converted the CSV event logs into XES format via ProM.

Once we obtained the event logs from the Patron dataset, we proceeded to the data mining and analysis stage.

3.5. Process data analysis and initial observations

By looking at the data through the lens of Definition 1–3, we could summarise its characteristics as shown in Tables 2 and 3.

The main log (labeled, GP16-20) covered 45 months. The GP16-20 log counts 2.5 million traces, of which 1.0 million (41.4%) are distinct. Most of the distinct traces are rarely observed. In all the logs (GP16-GP20), 98% of the distinct traces are observed between 1 and 5 times (see Table 3), and less than 0.03% of distinct traces are observed more than a thousand times. The traces include 31.8 million events, which, to the best of our knowledge, dwarf any of the real-life public logs used in automated process discovery research [3]. The trace length varies widely, with minimum, average, and maximum length of 1, 12, and 2317 events (respectively).

Table 2 provides summary statistics on the event logs.

| Log   | Total Traces | Distinct Traces | Distinct Events | Average Len. (events) | Max Len. (events) | Min Len. (events) |
|-------|--------------|-----------------|-----------------|-----------------------|------------------|------------------|
| GP16  | 6,976,703    | 6,976,323       | 31.8 million    | 6.9 million           | 2317             | 1                |
| GP17  | 6,291,366    | 6,291,086       | 1.1 million     | 5.1 million           | 12               | 1                |
| GP18  | 6,253,267    | 6,252,786       | 0.8 million     | 4.0 million           | 12               | 1                |
| GP19  | 6,212,224    | 6,211,744       | 1.3 million     | 5.0 million           | 12               | 1                |
| GP20  | 6,320,359    | 6,320,079       | 1.2 million     | 4.9 million           | 12               | 1                |

Given that our goal was to compare the patients behaviour in the months between March and November 2020 against the patients behaviour in the same timeframe of the past four years, we divided the log into five sublogs (namely, GP20, GP19, GP18, GP17, GP16), each of them capturing the 9-month timeframe in one of the five years under analysis. Such an approach is common for performing process behavioural comparison – known in the area of process mining as process variant analysis [4]. Looking at Tables 2 and 3, we notice that dividing the GP16-20 log into five sublogs does not affect much the variety of the process behaviour. Although the absolute number of events and traces reduces, each of the five (sub) logs maintains remarkable characteristics; i.e., 5.1 million (GP20 log) to 6.9 million (GP19) events, and 401 thousand (GP20 log) to 520 thousand (GP17 log) traces (on average, 41% distinct). As a comparison, the largest real-life event log used in the series of business process intelligence challenges had 1.6 million events. By analysing the characteristics of these five logs, we can immediately draw some initial observations.

Observation 1. In 2020, there was an average drop of 22.8% of patients accessing GP clinic healthcare services, compared to 2016–19. This is captured by the decrease of the total number of traces observed in the GP20 log, 401,370 as opposed to an average of 520,304 across the previous years – having min and max of 507,075 and 531,618.

Observation 2. In 2020, GP clinic healthcare processes maintained their high-level behavioural variety. This is captured by the almost constant percentage of distinct traces, 41.6% in 2020, and 41.4% on average over 2016–19 (meaning that each distinct healthcare process execution was observed approximately two times over 9 months in all the five years) but also, and perhaps most importantly, by the distribution of distinct traces over a set frequency buckets. Table 3 shows that in each year, the percentage of distinct traces that were observed between 1 and 5 times is constant at 98.2%.

Observation 1 was expected, given that a strict lockdown was enforced in Victoria from 16-March to 21-June and from 04-July to 09-Nov, that possibly deterred patients from accessing healthcare for what they considered minor issues. Observation 2 had a surprising nature, in fact, intuitively, we would expect that the combination of lockdown and pandemic would foster standardization in healthcare processes (i.e., less variability).

We explored the distribution of the activities over time, their frequencies, and how they varied over the five years. This information is shown in Fig. 1. Fig. 1a and b show the absolute and relative frequencies of each of the seven activities over the five years. The absolute frequency of an activity is the total number of times an activity is observed in all the traces recorded in the event log, while the relative frequency of an activity is the ratio of its absolute frequency over the total number of activities recorded in the event log – in this case, the total number of activities is equivalent to the total number of events recorded in the event log. Fig. 1c to i show the absolute frequency of each activity over time, month by month; and Fig. 1j shows the changes in the absolute frequency for each of the activities in 2020 when compared to the previous four years. In Fig. 1j, a negative change highlights a decrease in the absolute frequency of a specific activity in 2020 with respect to another year, while a positive change highlights an increase in the absolute frequency of a specific activity in 2020 with respect to another year. From the data captured in these plots, we can observe the following.

Observation 3. In 2020, the relative frequency of activity B (GP records a measurement) dropped to 9.0% from an average of 12.1%. Although this seems a small variation, we note that in 2019, 2018, 2017, and 2016, the relative frequency of activity B was remarkably stable at 12.2%, 12.0%, 12.1%, and 12.2% (respectively).

Observation 4. In 2020, the relative frequency of activity D (GP prescribes a refill) increased to 9.2% from an average of 5.0%. Also in this case, we note that in 2019, 2018, 2017, and 2016, the relative frequency of activity D was somewhat stable at 5.8%, 5.1%, 4.6%, and 4.4% (respectively).

Observation 5. In 2020, the change in the absolute frequency of activity G (vaccinations are administered/recorded) is remarkably low when compared to the change in the absolute frequency of other activities (e.g., activity A, B, and E, with an average decrease of 43.2% and 33.3%). In fact, the absolute frequency of activity G decreased of only 12.0% and 6.1% – compared to 2019 and 2018, and it increased of 7.4% and 16.9% – compared to 2017 and 2016. Furthermore, the absolute frequency of activity G is concentrated in the months of March and April, in contrast with the other years, where activity G is mostly observed in April and May.

Observation 3 can be straightforwardly interpreted. Given that activity B represents a GP taking and recording a measurement of the patient (e.g., measuring and recording the patient blood pressure), its decrease can relate to the actual implementation of safety measures – GP doctors may have avoided interacting with the patients unless strictly necessary.

Observation 4 represents an increase in medication refills. In particular, looking at Fig. 1f, which captures the activity D distribution over the nine months, we note a clear spike in March, April, June, July, and September. This can relate to an overstocking of drugs by patients that could not risk to run out of their medications. We remind that, during the early COVID-19 pandemic, overstocking was a phenomenon observed across a variety of products from food to toilet paper, known also as panic buying [47]. However, taking into account the changes of absolute frequency for activity D (see Fig. 1j), we can observe that drug prescriptions have increased steadily in the past four years with an average increase of 14.2%. Given that also a similar trend can be observed for activity C (capturing a first-time drug prescription), we cannot conclude that the increase observed in activity D derived exclusively from the COVID-19 pandemic context.

Lastly, Observation 5 is probably the most interesting one, also because it is in contrast with research findings of similar studies conducted in different geographical areas during the equivalent seasons [13,14]. The data clearly shows that vaccinations were not substantially
impacted in 2020, with a decrease in absolute frequency that is lower than the one of other activities (see Fig. 1J). Leaving aside medication-related activities (i.e., activities C and D), other activities reported an absolute frequency drop of between 23.9% (on average for activity A–GP Visit) and 43.2% (on average for activity B–measurement). While activity G (vaccinations) reported a maximum absolute frequency drop of 12.8% (compared to 2019) and an average drop of 1.3%. If we consider this in light of the total drop of the activities observed in 2020 (23.6% on average, see Table 2 – total events), the drop of vaccination activities is well below the average drop of the other medical activities. In addition, there is a noticeable shift in the vaccination timeline for the year 2020, bringing the vaccinations forward of one month. Observation 5 set a direction for additional analysis, which led us to additional findings that we will discuss in depth in Section 4.

3.6. Challenge 1 – Imprecise timestamps

Until now, we have described and analysed the data in general terms. Although we approached it from a process perspective, identifying the process activities and their execution over time, we have not discussed nor analysed the process behaviour, i.e., how such activities follow one another, and what their execution leads to. To analyse the process behaviour, process mining methodologies and tools often rely on directly-follows relations [48] (see Definition 4), especially, for automated discovery of process models [49], and for process variant analysis [27, 50, 28].

Recalling the event log definition (Definition 2), given that the order of the events in an event log is imposed by the order of their timestamps, incorrect or imprecise timestamps can have a significant (negative) impact on the identification of directly-follows relations and, consequently, on the output of process mining tools that rely on directly-follows relations. This is a well-known problem in the field of process mining [51, 52], especially in healthcare [16], where activities are documented manually. We recall that also in our case the event time is documented manually. We recall that also in our case the event time

3.7. Challenge 2 – Unbounded process instances

Once we solved the problem of imprecise timestamps, we focused on the process behaviour, analysing how the process activities follow one another and what their execution leads to. However, we note that our approach of fixing imprecise timestamps due to high-level granularity can be generalized to virtually any other context when the objective of the process analysis is the comparison of process variants, so it should not be considered as an ad-hoc approach for our specific scenario. However, we acknowledge that to define the logical order on the activities, the input of domain experts may be required. In our case, we relied on the experience in general practice medicine of the co-authors Dr. Capurro and Dr. Manski-Nankervis.

Lastly, we note that the time complexity of our approach is linear on the number of events contained in the event log, making it not only effective but also efficient.
of a trace in the input event log to be the start of the process instance, and the last event of a trace to be the end of the process instance. We can, however, identify the most appropriate start and end events given a process instance. This can be achieved by narrowing down the scope of an unbounded process instance, for example, by focusing on a single GP visit or a single health issue/procedure. To do that we devised an algorithm that leverages domain experts’ knowledge, once again, the co-authors Dr. Capurro and Dr. Manski-Nankervis.

We started from the assumption that a process instance should begin with a visit to the GP doctor (i.e., activity A), effectively making activity A the only possible start event of a trace. Any subsequent activity different than activity A (i.e., activities B to G) is assumed to be the follow-up of the initial visit to the GP doctor. However, when a second activity A is observed for the same process instance, we have to distinguish two cases: i) the new activity A is a follow-up of the past activities; ii) the new activity A is not related to the past activities (i.e., this would trigger a new process instance). We distinguished the two cases on a time basis. Precisely, if the new activity A is more than six months away from the first observed activity A and more than one month away from the last observed activity of the current process instance, we are in case ii); otherwise, we are in case i). These time thresholds were set empirically following the domain experts.

Algorithm 1 describes a generalisation of our approach to generate traces from a given event log containing unbounded process instances. The algorithm takes in input the log (L), a set of allowed start activities (α) – in our case containing only activity A, and two time thresholds Δ₀ and Δₐ – in our case six- and one-month respectively. Three data structures are initialised (see lines 1 to 3): i) a map linking an event ID to its trace (Π) – representing the collection of traces to output; ii) a map linking an event ID to the timestamp of the first event in the corresponding trace (T₀); and iii) a map linking an event ID to the timestamp of the last observed event in the corresponding trace (Tₐ). Then, we read the log (L) one event at a time, starting from its first event (e, line 4).

If the event ID (eₚ) is not yet in the map Π and the event activity (eₚ) is in the set α, we create a new empty trace (τ), we append e to τ, we add the event ID and the trace to the map Π, we save the timestamp of e in T₀ and in Tₐ (lines 5 to 10).

If eₚ is already mapped in Π and eₚ is not an allowed start activity (line 12), we retrieve the trace linked to the event ID (Π(eₚ)) and we append e to that trace (line 13). Then, we update the timestamp information by overwriting the last observed event timestamp in the map Tₐ (line 14).

If eₚ is already mapped in Π and eₚ is an allowed start activity (line 12), we distinguish the two possible cases mentioned above. Case ii), if eₚ is less than or equal to Δ₀ or less than or equal to Δₐ, then we append e to the already existing trace Π(eₚ) (as just described above – see lines 16 to 18). Otherwise, Case i), we create a new event ID (that is not present in the event log), we link the new event ID to the existing trace in Π that is mapped to eₚ, we create a new empty trace (τ), we append e to τ, we add the event ID and the trace to the map Π, we save the timestamp of e in T₀ and in Tₐ (lines 21 to 20).

Once all the events in the event log have been read, Algorithm 1 returns the map of event IDs and the corresponding traces.

Assuming that accessing the maps is a constant-time operation, as it is the case in modern object-oriented programming languages, Algorithm 1 has a linear time complexity on the number of events contained on the event log.

We note that the information shown in Table 2 is the one obtained after the execution of Algorithm 1. The column filtered events reports the number of events that were removed by applying Algorithm 1, i.e., events that are not preceded by an activity A. On average, we removed 3.6% of events from the data, which is a negligible amount.

Algorithm 1. Generate traces from the event log

---

7 This can be achieved by manipulating the current eₚ.
3.8. Challenge 3 – Any process behaviour is allowed

At this stage, we can finally turn our attention to the process behaviour analysis, by leveraging process mining techniques [46]. Since we are interested in identifying process behavioural differences over five different timeframes (each captured in an event log), the appropriate process mining techniques are in the class of automated process discovery [3] and process variant analysis [4]. Automated process discovery techniques receive in input an event log and automatically produce a process model, which is a graphical representation of the process behaviour, such as a workflow chart, a Petri net, or a BPMN model. By looking at different process models, it is possible to detect behavioural differences. On the other hand, process variant analysis techniques receive two event logs and automatically produce a model which is a graphical representation of the process behaviour, such as a workflow chart, a Petri net, or a BPMN model. From both classes of techniques, we selected three state-of-the-art tools, based on previous studies evaluations [3,28,29] which are: Fodina [21], Inductive Miner [23], Split Miner [22] (for automated process discovery); and process comparator [27], fingerprints-based variant analysis [28], and variant analysis via declarative rules [29] (for variant analysis).

We ran these six techniques on the Virtual Machine hosting the data, which was equipped with a Xeon(R) CPU E5-4620 v2 @2.60 GHz with 8 virtual processors and 64 GB of RAM. All the techniques have a Java implementation either as a standalone tools [22,29] or as a ProM plugins [23,21,27], except the fingerprints-based variant analysis [28] whose implementation is on Python. The Java-based tools (including ProM plugins) were provided with 40 GB of heap space. For executing our tests, we were forced to use a 2-h timeout, this was a security measure of the Virtual Machine for running applications via command-line. However, we note that is a very generous timeout according to recent benchmarks and evaluations [3,57,58,29]. Lastly, all the tools were run with their default parameters, in particular, their default filtering thresholds.

First, we attempted to discover a process model from each of the five event logs, by running each of the three automated process discovery tools. Fodina was not able to output a sound process model from any of the five event logs within the timeout. Inductive Miner discovered exactly the same process model from all the five logs, which is the one capture in Fig. 2a. Split Miner discovered the same process model from four logs (GP16-GP19, Fig. 2b), but a different one from the GP20 log (Fig. 2c).

The process models shown in these figures are modelled using the BPMN model.\(^8\) By looking at different process models, it is possible to detect behavioural differences.
BPMN 2.0 standard notation. For the readers who are unfamiliar with this notation, we remind that: the round graphical elements with thin and thick border represent respectively the start and the end events of the process; the squared boxes represent the activities that can be executed during a process instance; a squared box marked with the symbol \( \text{arr} \times \text{col} \) represent an activity that can be consecutively executed several times, before allowing the process execution to continue; the diamonds with an \( X \) represent decision points (when having multiple outgoing arcs) or passive join points (when having multiple incoming arcs); lastly each arc in a BPMN 2.0 model captures the allowed process execution flow.

The process models we discovered are not structurally complex (e.g., spaghetti-models), but they clearly show that a large variety of behaviour is allowed – with few constraints and many cyclical patterns. This finding highlights that the GP day-to-day healthcare processes have a behavioural degree of freedom that is not comparable with most business processes, allowing a vast amount of different behaviour to be executed and repeated over time.

For instance, considering the model discovered by Inductive Miner from any of the five logs (Fig. 2a), we note that all the seven activities can be executed in parallel in any order, all the activities can be skipped except activity \( A \), and four of them can be repeated any number of times (\( A, B, E, F \)). Although the model over-generalize the behaviour recorded in the log, it gives a clear idea of the degree of freedom of the process under analysis. As a comparison, if we consider the model discovered by Split Miner from the GP20 log (Fig. 2c), the process model highlights that activity \( A \) must be the first to be executed, but it can be repeated after executing other activities. The model correctly captures other behavioural constraints, for example, activity \( C \) must precede activity \( D \), and activity \( G \) can follow only after activity \( D \). However, the model allows all the activities to be repeated any number of times. The difference between the model discovered by Split Miner from the GP20 log and any of those discovered from the GP16-GP19 logs is minimal, and mostly involving activity \( G \). The latter is not anymore following activity \( D \).

From this automated process discovery exercise, we can see that the models discovered by Inductive Miner and Split Miner highlight no substantial difference over the four years preceding 2020, suggesting a broadly in-line process behaviour from 2016 to 2019. However, the model discovered by Split Miner from the GP20 catches one small difference. To better investigate such potential difference(s), we continued our analysis applying the process variant analysis tools.

We ran each of the three tools four times, providing in input four pairs of logs: (GP16, GP20); (GP17, GP20); (GP18, GP20); (GP19, GP20). By doing so, each tool would identify the differences in the observed process behaviour between each pair of years, perfectly addressing our research and analysis goal.

Also in this case, one of the three tools \([28]\) was unable to provide an output within the timeout, or it was unable to identify any statistically significant difference.

The tool of Bolt et al. \([27]\), namely \textit{process comparator}, identified some differences. The output of the comparisons (GP18, GP20) and (GP19, GP20) are shown in Fig. 3. In Figs. 3a and b, a blue arc between two nodes (which represent activities) highlights that the waiting time between the completion of one activity (the arc’s source) and the other (the arc’s target) is significantly greater in the log GP20, instead, when the arc is red the processing time is significantly greater in the comparing log – respectively GP18 and GP19. The different shades of blue and red provide an intuitive idea of the magnitude of the difference (the darker the shade the greater the difference), but not an exact estimate. An identical decoding reasoning applies for interpreting the color-coding of Figs. 3c and d, with the only difference being the comparison metric, which is not the waiting time but the frequency of activities (i.e., color-coded nodes) and the frequency of the directly-follow relations (i.e., color-coded arcs).

The process comparator was able to identify statistically significant differences (\( p < 0.05 \)) both in terms of time elapsed between two consecutive activities and the frequency of observing two consecutive activities. We remind that the process comparator does not a direct difference between metrics, but analyses and contextualize them within the respective logs. Considering some of the interesting differences identified by the process comparator, we can see that both in 2018 and 2019 the time elapsed between a GP visit (\( A \)) and a vaccination (\( G \)) was significantly greater than in 2020. On the other hand, the time elapsed between the prescription of a medical test (\( E \)) and its results (\( F \)) was greater in 2020 than in 2018 and 2019. When we look at differences in terms of frequency, we can see that many pair of activities were more

![Fig. 3](https://www.bpmn.org/)

**Fig. 3.** Variant analysis output using process comparator \([27]\), the color-coding is explained in the text.

---

**Table 4**

Top-10 differences between healthcare processes in 2020 (Variant X) and 2019 (Variant Y) \([29]\).

| Rank | Difference | Induced by | Change in Probability | Probability in 2020 | Probability in 2019 |
|------|------------|------------|-----------------------|---------------------|---------------------|
| 1    | In Variant X it is 16.7% more likely than Variant Y that if [B] occurs, also [D] occurs | In Variant Y it is 10.5% more likely than Variant X that if [A] occurs, also [B] occurs | 0.056 | 0.039 |
| 2    | In Variant X it is 16.6% more likely than Variant Y that if [E] occurs, also [D] occurs | In Variant X it is 11.9% more likely than Variant Y that if [A] occurs, also [D] occurs | 0.055 | 0.036 |
| 3    | In Variant X it is 16.3% more likely than Variant Y that if [F] occurs, also [D] occurs | In Variant X it is 11.4% more likely than Variant Y that if [F] occurs, also [B] occurs | 0.055 | 0.036 |
| 4    | In Variant X it is 11.9% more likely than Variant Y that if [A] occurs, also [D] occurs | In Variant X it is 11.4% more likely than Variant Y that if [F] occurs, also [B] occurs | 0.055 | 0.036 |
| 5    | In Variant X it is 11.4% more likely than Variant Y that if [F] occurs, also [B] occurs | In Variant X it is 11.3% more likely than Variant Y that if [D] occurs, also [C] occurs | 0.055 | 0.036 |
| 6    | In Variant Y it is 11.3% more likely than Variant X that if [D] occurs, also [C] occurs | In Variant Y it is 11.1% more likely than Variant X that if [E] occurs, also [B] occurs | 0.055 | 0.036 |
| 7    | In Variant Y it is 11.1% more likely than Variant X that if [E] occurs, also [B] occurs | In Variant X it is 10.7% more likely than Variant Y that if [D] occurs, also [C] occurs | 0.055 | 0.036 |
| 8    | In Variant Y it is 10.7% more likely than Variant X that if [D] occurs in a process instance | In Variant X it is 10.7% more likely than Variant Y that if [C] occurs, also [D] occurs | 0.055 | 0.036 |
| 9    | In Variant Y it is 10.7% more likely than Variant X that if [C] occurs, also [D] occurs | In Variant Y it is 10.5% more likely than Variant X that if [A] occurs, also [B] occurs | 0.055 | 0.036 |

10 https://www.bpmn.org/.
Table 5
DFG Matrix of the GP20 log (Inc. relative frequencies of the directly-follows relations in brackets).

| Activity | A     | B     | C     | D     | E     | F     | G     |
|----------|-------|-------|-------|-------|-------|-------|-------|
| From     |       |       |       |       |       |       |       |
| A        | 916053 (20.3%) | 202147 (4.5%) | 263395 (5.8%) | 168861 (3.7%) | 179713 (4.0%) | 137296 (3.0%) | 111189 (2.5%) |
| B        | 108522 (2.4%) | 60445 (1.3%) | 33589 (0.7%) | 16939 (0.4%) | 20640 (0.5%) | 157124 (3.5%) | 10187 (0.2%) |
| C        | 29990 (0.7%) | 3901 (0.1%) | 64 (0.0%) | 257958 (5.7%) | 9254 (0.2%) | 3929 (0.1%) | 2196 (0.0%) |
| D        | 217229 (4.8%) | 29634 (0.7%) | 406 (0.0%) | 972 (0.0%) | 76571 (1.7%) | 30394 (0.7%) | 16982 (0.4%) |
| E        | 126833 (2.8%) | 4963 (0.1%) | 2343 (0.1%) | 1589 (0.0%) | 8439 (0.2%) | 222635 (4.9%) | 11616 (0.3%) |
| F        | 298881 (6.6%) | 87045 (1.9%) | 9573 (0.2%) | 26567 (0.4%) | 3681 (0.1%) | 4079 (0.1%) | 2143 (0.0%) |
| G        | 89601 (2.0%) | 18193 (0.4%) | 102 (0.0%) | 152 (0.0%) | 3750 (0.1%) | 12611 (0.3%) | 99 (0.0%) |

Table 6
DFG Matrix of the GP19 log (Inc. relative frequencies of the directly-follows relations in brackets).

| Activity | A     | B     | C     | D     | E     | F     | G     |
|----------|-------|-------|-------|-------|-------|-------|-------|
| From     |       |       |       |       |       |       |       |
| A        | 1190430 (19.5%) | 449745 (7.4%) | 299143 (4.9%) | 70762 (1.2%) | 337449 (5.5%) | 122043 (2.0%) | 121266 (2.0%) |
| B        | 219968 (3.6%) | 107579 (1.8%) | 78582 (1.3%) | 13069 (0.2%) | 74732 (1.2%) | 221429 (3.6%) | 12921 (0.2%) |
| C        | 46794 (0.8%) | 11415 (0.2%) | 39 (0.0%) | 294829 (4.8%) | 26567 (0.4%) | 3681 (0.1%) | 4079 (0.1%) |
| D        | 179501 (2.9%) | 25863 (0.4%) | 145 (0.0%) | 456 (0.0%) | 88901 (1.5%) | 11993 (0.2%) | 15318 (0.3%) |
| E        | 109522 (2.4%) | 60445 (1.3%) | 33589 (0.7%) | 16939 (0.4%) | 20640 (0.5%) | 157124 (3.5%) | 10187 (0.2%) |
| F        | 126835 (2.8%) | 40963 (0.9%) | 2343 (0.1%) | 1589 (0.0%) | 8439 (0.2%) | 222635 (4.9%) | 11616 (0.3%) |
| G        | 298881 (6.6%) | 87045 (1.9%) | 9573 (0.2%) | 4856 (0.1%) | 124704 (2.8%) | 376932 (8.4%) | 2143 (0.0%) |

frequent in 2020 than in 2019 and 2018, for instance: (A; G); (A; C); (A; D); (C; D). Interestingly, the pair (A; B) was observed much less in 2020 than in the previous two years. While these differences support Observations 3–5 (see Section 3.5), the process comparator detected many other differences that we could not justify or link to our initial observations. Once again, this can be related to the amount and variability of the behaviour recorded in the event logs.

The output of the process comparator requires some interpretation from the analyst, which sometimes is not straightforward. Instead, the tool of Cecconi et al. [29] provides natural language descriptions of the identified differences. The top-10 differences identified by this tool [29] and in the GP20 and GP19 (Table 4) highlight the decrease in the number of observations B in the healthcare processes of 2020 (see Table 4, rows 5, 7 and 10), and an increase in the number of observations of activity D in the healthcare processes of 2020 (see Table 4, rows 1–4 and row 9). Similar results were obtained when comparing the data from the 2018 against the data from the 2017, 2016.

At this point, we took a step back, and decided to review the behaviour recorded in each of the five event logs by visualising their directly-follows graphs. For reasons of space, clarity, and simplicity, we report the DFG of only two event logs (GP20 and GP19) and in their matrix form, where each matrix row (and column) represents a node of the DFG – i.e., an activity; and each cell of the matrix captures the frequency of the edge between the two nodes – i.e., how many times we observe in the event log a directly-follows relation between two activities.

Tables 5 and 6 report the DFGs in matrix form of the event logs GP20 and GP19 (respectively). For clarity, in each DFG matrix cell, we have also reported in brackets the relative frequency of the directly-follows relation (i.e., the ratio of the number of times we observe a directly-follows relation over the total number of directly-follows relations observed in an event log). We note that any directly-follows relation can be observed in the two DFGs. Although some of them are rare (e.g., C→C, with a frequency in the order of hundreds), the vast majority can be observed in the two DFGs.
be observed with a frequency in the order of thousands – although they represent a tiny percentage of the total number of the observed directly-follows relations. According to this process data, the GP day-to-day healthcare process allows for executing medical activities in any order. Consequently, the data generated from such kind of process is very complex and challenging to analyse with state-of-the-art process mining techniques – as we have seen with Fodina [21] and Taymouri’s tool [28], both unable to output sound results. An alternative would be to remove some behaviour by applying filtering techniques [59–61], however, we recall that all the automated process discovery algorithms that we used already apply a filter [21,23,22], as well as two of the three variant analysis approaches [27,29].

While in a business context some infrequent behaviour may be a violation of compliance or internal business rules, in our context, all Top-20 traces, extract of the traces including activity G. Table 8

| Variant analysis approaches 
|---|
| that we used already apply a filter [21,23,22], as well as two of the three variant analysis approaches [27,29].

With that in mind, we considered only the most frequent behaviour. Table 7 reports the top-20 most frequent traces that we could observe in each of the five event logs. Scanning carefully through Table 7, we notice that in 2020, traces containing the activity G were more frequent than other years (for clarity, we reported these traces in Table 8). In 2020, not only the traces containing the activity G were more frequent, but they accounted for the 25% percent of the most frequent behaviour (5 traces out of 20). This finding is remarkable, and when paired it with Observation 5 (discussed in Section 3.5) clearly hints to a variation in the behaviour involving vaccinations during the 2020. A similar reasoning can also be done for traces containing the activity D (see A, D), across the five logs. Further investigation of the most frequent process behaviour may reveal several additional differences, but within the scope of this study, we decided to investigate the specific behavioural difference related to activity G and its traces among the top-20.

Our process mining analysis highlights two limitations of the state-of-the-art process mining techniques that we used in this study:

1. Process mining techniques for automated process discovery and process variant analysis suffer of scalability and/or quality issues when they deal with too much and too variable behaviour.
2. Automated process discovery techniques try to capture as much behaviour as possible from the event log, filtering infrequent behaviour only when it is strictly necessary to either simplify the process model or increase its accuracy. However, depending on the context, one may be interested in capturing very little process behaviour from the event log – requiring special filtering techniques.

While limitation 1 is ground for future research directions and studies. Limitation 2, at the moment, can be addressed manually, by applying ad-hoc filters of the process behaviour recorded in the event logs (as we did). The best ad-hoc filters must be identified by domain experts, often on a trial-and-error basis, and applied either via ProM plugins or commercial tools such as Celonis, Disco, or Apromore (which we used). Future process mining techniques should allow the user to automatically design such filters, without relying on domain experts knowledge. For instance, by automatically analysing the outputs of a set of process mining techniques (e.g., both automated process discovery and variant analysis) – as we did manually.

Once we narrowed our attention down to only the most frequent traces containing a vaccination event (activity G), we could easily discover a clear and simple process model for each of the five years and identify the differences, Figs. 4a to c show the process models. We note that in 2016, 2017, and 2018, the most frequent vaccination process is the same (Fig. 4a), changes in this process are minimal in 2019 (Fig. 4b), and substantial in 2020 (Fig. 4c). In fact, comparing to years 2016–2018, in 2019 the only difference is that in 1,414 process executions (out of 14,792, i.e., 9.6% of the times), a vaccination (activity G) would be administered after two visits to the GP (activity A). While, in 2020, we can frequently observe an activity B right after the vaccination (8.0% of times), two visits to the GP before the vaccination (6.7% of times), and two visits to the GP after the vaccination (5.3% of times).

Lastly, we analysed each of the process traces by looking into the distribution of the executed activities over time, we reported this information in Figs. 5a to f. At this point, it is evident that the differences in behaviour were not only in terms of how the activities were executed (i.e., their order and frequencies) but also when. We summarised these findings in the following observation.

Observation 6. In 2020, we can observe a clear (left-) shift (i.e., towards March) and early peak in the distribution of the activities executed within the most frequent behaviour of the vaccination process, as well as a different trend when compared to the past four years, which holds for all the activities involved in the vaccination process. Furthermore, the most frequent vaccination process in 2020 was more complex than the previous four years, allowing for more behavioural variants with frequently requiring additional activities (activity A, and B).

The observed change can be explained by the recommendation that

---

### Table 8

| Top-20 traces, extract of the traces including activity G. |
|---|
| 2020 | 2019 | 2018 | 2017 | 2016 |
| A.G | 15619 | A.G | 9145 | A.G | 6849 | A.G | 6549 |
| A.G.A | 10885 | A.G.A | 2295 | A.G.A | 1791 | A.G.A | 1743 |
| A.G | 3348 | A.G.A | 2295 | A.G.A | 1791 | A.G.A | 1743 |
| A.G.A | 2493 | A.G.A | 2295 | A.G.A | 1791 | A.G.A | 1743 |
| A.A.G | 1414 | A.A.G | 1414 | A.A.G | 1414 | A.A.G | 1414 |
| A.G.A | 1578 | A.G.A | 1578 | A.G.A | 1578 | A.G.A | 1578 |
| A.G.A.A | 1267 | A.G.A.A | 1267 | A.G.A.A | 1267 | A.G.A.A | 1267 |

---

**Fig. 4.** Vaccination process models (most frequent behaviour only).
Australians receive their influenza vaccinations before the normal season (April-May). This recommendation was broadly advertised to minimize a possible double hit to the healthcare system: an epidemic of SARS-CoV-19 in addition to the usual Fall/Winter influenza season. In the next section, we will discuss more in depth this observation from a medical perspective.

3.9. From process to data mining

Our process mining analysis, including the extraction of process data, its plotting (e.g., Fig. 1), and the application of state-of-the-art process discovery and process variant analysis tools, provided us with a clear picture of the GP day-to-day healthcare process as well as a lead to follow. It allowed us to identify relevant differences in the behaviour of the patients in 2020, in particular, when considering the vaccination process. Accordingly, we refined our original research question into two sub-questions. RQ1. What type of vaccines have driven the frequency increase in the most frequent vaccination process traces? RQ2. What are the differences between vaccination behaviour of different age classes, i.e., children (0–17 years), adults (18–64 years), and elderly people (65+)?

To answer these research questions, process mining techniques can provide little help. Process mining and, more in general, process science and process thinking can be a lighthouse in a ocean of data. However, it is difficult to dig deeper by only relying on process mining techniques, given that at the current stage they do not take into account rich perspectives surrounding the process behaviour. In fact, to the best of our knowledge, there are no reliable and effective process mining techniques – in the area of automated process discovery [3] and process variant analysis [4] – that give a global picture of the process, taking into account all the additional data recorded in the event attributes available in the event log. For instance, to answer our research questions, the crucial event attributes are patient age and vaccine type, but the integration of this information in a process model is not a trivial problem. Besides, we note that our refined research questions are data mining oriented. In fact, process mining and data mining are complimentary, and future research directions should leverage this relation between the two disciplines to bring them together.

To continue with our analysis, we extracted all the data regarding vaccination events (activity G) from the original dataset (GP16-20 event log), and we analysed the different vaccine types and the immunity they provide. Table 9 shows a mapping between the vaccines and the labels we will use to simplify the presentation of the data.

![Fig. 5. Vaccination process (most frequent behaviour only), breakdown by activities. Each plot reports on the y-axis how many times we observed an activity of the process in a given fortnight (x-axis, from the 5th to the 23rd fortnight of a year). The plots/activities are ordered following the behaviour represented by the process models in Fig. 4, the order of the activities is captured in brackets. Activities that are optional (i.e., may not be executed, according to the process behaviour) are explicitly mentioned.]

| Table 9 Encoding of vaccines. |
|-----------------------------|-----------------------------|
| Label | Provided Immunity |
| V1   | Cholera          |
| V2   | Coxiella Burnetti|
| V3   | Diphtheria       |
| V4   | Haemophilus B    |
| V5   | Hepatitis A      |
| V6   | Hepatitis B      |
| V7   | HPV              |
| V8   | Influenza        |
| V9   | Japanese Encephalitis |
| V10  | Measles          |
| V11  | Meningococcal    |
| V13  | Pneumococcus     |
| V14  | Poliomyelitis    |
| V15  | Rabies           |
| V16  | Rotavirus        |
| V17  | Salmonella typhi |
| V18  | Tetanus          |
| V19  | Tuberculosis     |
| V20  | Varicella Zoster |
| V21  | Yellow Fever     |

11 https://www.abc.net.au/news/health/2020-04-01/australians-urged-to-get-flu-vaccination-coronavirus-covid-19/12107264.
Observation 7. In 2020, there was a surge of influenza (V8) and pneumococcus (V13) vaccinations (see Fig. 6, vaccine V8 and V13), predominant in adults and elderly people, and in contrast with a decrease of these vaccinations for young people (see Figs. 6b to e, V8 and V13). The increase is even more startling when we consider that all the other vaccines suffered a decrease of approximately 50% (on average).

Furthermore, we were able to observe that vaccines associated with international travel requirements (i.e., Yellow Fever, Japanese Encephalitis) practically disappeared, probably a result of international border closures. Lastly, it is worth mentioning that by looking at Fig. 6d also the vaccinations for Rotavirus (V16) and Tuberculosis (V19) surged in 2020, however, this change turns immaterial when looking at the absolute number of these vaccinations (Fig. 6a) which are in the single digit order – e.g., in 2020 there were 2 Rotavirus vaccinations while in 2017 only 1.

In the next section, we explore more in depth these implications from a medical point of view.

3.10. Limitations of the study

In this section, we described how we have analysed the data and the observations we could draw from it. To analyse the data, we followed a
well-known methodology, PM² [46]. We note that the latter was designed to be applied in a business-context. However, we argue that this does not pose a threat to its applicability in healthcare, in fact, we were able to adhere to its stages from start to end, with the exception of omitting the execution of the process improvement stage, since it was out of the scope of this study.

As broadly discussed in this section, the data we had access to, i.e., the Patron dataset, was far from ready for a process mining analysis. Not only the data was heterogeneous both within a given year and across different years, but its structural quality did not meet the minimum requirements for applying and running state-of-the-art process mining tools, for instance, the timestamps had a day-granularity rather than hour and minute. To overcome the data quality issues, we proposed and applied two novel solutions. While these solutions allowed us to proceed further with our analysis, we recognise that different solution may have been designed which may have, in theory, led to different outcomes. Hence, the observations and the justifications we provide in this study should be interpreted in light of the procedures we applied from the data cleaning and filtering to the process mining analysis.

To execute the process mining analysis, we relied on a subset of the existing state-of-the-art process mining techniques for automated process discovery and variant analysis, which we selected according to the findings of the most recent literature reviews. While, in theory, applying other techniques may have yielded different or better results, we recall that the process mining techniques we used were the latest and the most reliable.

The observations reported in this study cannot be generalised to Australia, nor the state of Victoria. However, we note that the analysed data captured the behaviour of approximately 400 thousand patients (per year), which account for almost 6% of the entire population of the state of Victoria – a remarkable percentage, especially when we consider that not the whole population regularly visit GP clinics. While the observations reported in this study are derived from the data and, hence, objective, their analysis and our discussion represent our interpretation. We note that when providing an explanation for a specific observation we considered findings of other similar studies and the experience of two domain experts who co-authored this study (Dr. Capurro and Dr. Manski-Nankervis). In theory, alternative interpretations for some of our observations may be possible but, to the best of our knowledge, the one we provided in this study are the most reasonable and realistic.

4. Discussion

The study presented here represents the first use of process mining techniques to analyze the impact of the COVID-19 pandemic in health services utilization patterns in primary care. Using a combination of process mining techniques we were able to highlight several relevant changes in health services utilization patterns associated with the disruptions seen in 2020. In addition to these, we were able to highlight some limitations of the process mining tools available today, in particular, when applying them to analyze healthcare process data.

Overall, we observed a widespread reduction of GP activities during the period included in our study, when compared to the same period in the four preceding years, concordant with what has been reported in other countries [62]. It is expected that such a reduction of GP activities led to a reduction of specialists visits –given that the Australian is a referral-based system. The consequences of such additional potential reduction of healthcare activities remain to be seen. From the process perspective, in such a situation, we would have expected a reduction in the number of distinct healthcare process execution, instead, the degree of variety of process behaviour remained almost unchanged during this period – with 98% of distinct healthcare process executions observed only between 1 and 5 times.

One activity that showed a different behavior were drug prescriptions. We observed an increase in drug refill prescriptions, with peaks in March, April, July and September. These peaks are associated with periods immediately before lock-downs and might represent overstocking of chronic medications. This observation is in line with what has been observed in Australian national drug prescription databases [63].

The most notable changes were observed in activities involving vaccinations. First, we see that although there still was a reduction in the total number of vaccinations, the drop was relatively minor compared to the rest of the GP activities. Vaccinations dropped an average of 1.3% and all other activities dropped an average of 23.6%. This contrasts to what has been reported elsewhere, where the 2020 pandemic has been associated to significant reduction in vaccination rates [15]. When we look into specific vaccines, we can see an increase in influenza vaccinations together with an earlier peak. This is in line with public health campaigns urging citizens to get their annual influenza vaccines and prevent a double epidemic. Interestingly, in older adults we can see a parallel increase in pneumococcal vaccinations. The most likely explanation was the drive to reduce any preventable respiratory infection in preparation of the impending pandemic. Finally, vaccines normally recommended for international travel (Yellow Fever, Japanese Encephalitis, Cholera) practically disappeared, as a consequence of the severe limitations to international travel.

From the process mining perspective we faced several challenges related to the problem of analysing a vast amount of process execution data. To the best of our knowledge, the event log analysed in this study represents the largest real-life event log used for automated process discovery and process variant analysis, especially, in the healthcare context. We showed that traditional process mining tools present some limitations when attempting to analyze processes with high behavioral variability.

The first challenge consisted of imprecise timestamps, since the time granularity was limited to day-level, a recurrent problem in the healthcare context that yet has to be solved. In our case, we relied on clinical knowledge to address this issue by defining a sequence of clinically meaningful activities as a tie-breaker for activities that had identical timestamps.

The second challenge involved the identification of start and end events for a process that is, by nature, unbounded. Once again, we relied on domain expertise to overcome this problem and we presented a generalisation of our solution, suitable for various contexts, in the algorithm described in Section 3.

The third challenge was the amount of data itself, and its high behavioral variability, which disarmed state-of-the-art process mining techniques for automated process discovery and process variant analysis. Although the scope of this study was not to devise novel variants of these techniques to deal with such type of data, we highlighted possible directions for future research addressing the improvement of these techniques. Lastly, this study reminds us that state-of-the-art process mining techniques (such as those we used [21,23,22,27–29]) do not yet automatically analyse the event log information that is not related to the process behaviour and control-flow (e.g., patient age, medications, etc). This is a straightforward consequence of the existing process mining algorithms design, which do not take into account the additional information even when available. While first steps have been made towards the next generation of process mining algorithms [64], robust solutions have yet to be proposed. The existing limitation requires process analysts to integrate the process mining analysis with a data mining analysis. While this problem could be solved by further analysing the data from a different perspective, we call for future analysis methodologies and tools that automatically integrate both process and data perspectives.

5. Conclusion

This study represents the first application of process mining techniques to analyze the impacts of the COVID-19 pandemic in the patterns
of primary care service utilization, specifically, in the General Practice day-to-day healthcare processes of Victorian patients. Our analysis identified several relevant changes in the behavioural patterns of the patients. While some of these changes were expected, i.e., overall reduction in number of attended GP visits, some were not, i.e., increase in the number of medication prescriptions, less than expected drop in vaccinations, and increase of influenza and pneumococcus vaccinations – in contrast with research findings from different geographical areas.

The size of the data-set under analysis – counting 31-million events – and the variability of the observed process behavior were unique, and the challenges we faced and overcame during the process mining analysis clearly highlighted the need for improving existing process mining techniques, drawing directions for future work. In particular, future process discovery techniques should integrate in the discovered process models also data surrounding the process behaviour and its control flow. In the healthcare context, this data is the information capturing a patient profile (e.g., age, gender, etc) and their medical procedure (e.g., type of vaccination or prescribed medication). Furthermore, existing process mining techniques are not tailored to deal with the large amount of data that captures highly variable behaviour. Future research should consider the design of methods that can automatically filter process execution data to detect and extract the most relevant/interesting process behaviour (not necessarily the most frequent) by analysing the outputs of a set of process mining techniques (e.g., a combination of automated process discovery and process variant analysis). Lastly, as process mining applicability in the healthcare context gains momentum, novel process mining techniques should be tailored for such a context and leverage domain expertise to increase their effectiveness.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] W. van der Aalst, Process Mining - Data Science in Action, Springer, 2016.
[2] M. Dumas, M. La Rosa, J. Mendling, H.A. Reijers, Fundamentals of business process management, 2nd ed., Springer, 2018.
[3] A. Augusto, R. Conforti, M. Dumas, M. La Rosa, F. Maggi, A. Marrella, M. Meccia, A. Soo, Automated discovery of process models from event logs: Review and benchmark, IEEE TKDE 31(4).
[4] F. Taymouri, M. La Rosa, M. Dumas, F.M. Maggi, Business process variant analysis: Survey and classification, Knowl.-Based Syst. 211 (2021) 106557.
[5] S. Dunner, M. Stierle, M. Matzner, S. Baier, Conformance checking: a state-of-the-art literature review, in: Proceedings of the 11th international conference on subject-oriented business process management, 2019, pp. 1–10.
[6] L. Verenich, M. Dumas, M.L. Rosa, F.M. Maggi, I. Teinemaa, Survey and cross-benchmark comparison of remaining time prediction methods in business process monitoring, ACM Trans. Intell. Syst. Technol. (TIST) 10 (4) (2019) 1–34.
[7] E. Rojas, J. Munoz-Gama, M. Sepúlveda, D. Capurro, Process mining in healthcare: A literature review, J. Biomed. Inform. 61 (2016) 224–236.
[8] G. Leonardi, M. Strian, S. Guagliardi, A. Cavallini, S. Montali, Leveraging semantic labels for multi-level abstraction in medical process mining and trace comparison, J. Biomed. Inform. 83 (2018) 10–24.
[9] C. Alvarez, E. Rojas, M. Arias, J. Munoz-Gama, M. Sepulveda, V. Herskovic, D. Capurro, Discovering role interaction models in the emergency room using process mining, J. Biomed. Inform. 78 (2018) 60–77.
[10] J. Chen, L. Sun, C. Guo, W. Wei, Y. Xie, A data-driven framework of typical treatment process extraction and evaluation, J. Biomed. Inform. 83 (2018) 178–195.
[11] S. Yang, A. Sarcevic, R.A. Farneth, S. Chen, O.Z. Ahmed, I. Marsic, R.S. Burd, An approach to automatic process deviation detection in a time-critical clinical process, J. Biomed. Inform. 85 (2018) 155–167.
[12] N. Martin, J. De Weerdt, C. Fernández-Llatas, A. Gal, R. Gatta, G. Iba nez, O. Johnson, F. Manshardt, L. Marco-Ruiz, S. Mertens, et al., Recommendations for enhancing the usability and understandability of process mining in healthcare, Artif. Intell. Med. 109 (2020) 101962.

12 Victoria, Australia
[41] Cancer Australia, Review of the impact of covid-19 on medical services and procedures in australia utilising mbs data: Skin, breast and colorectal cancers, and telehealth services (2020).
[42] T. Robertson, E.D. Carter, V.B. Chou, A.R. Stegmuller, B.D. Jackson, Y. Tam, T. Sawadogo-Lewis, N. Walker, Early estimates of the indirect effects of the covid-19 pandemic on maternal and child mortality in low-income and middle-income countries: a modelling study, Lancet Global Health 8 (7) (2020) e901–e908.
[43] Special feature: immunization and covid-19 (Jul 2020). URL: https://www.who.int/immunization/monitoring_surveillance/immunization-and-covid-19/en/.
[44] Data for Decisions and the Patron Program. URL: https://medicine.unimelb.edu.au/school-structure/general-practice/engagement/data-for-decisions.
[45] R. Canaway, D.I. Boyle, J.-A.E. Manski-Nankervis, J. Bell, J.S. Hocking, K. Clarke, M. Clark, J.M. Gunn, J.D. Emery, Gathering data for decisions: best practice use of primary care electronic records for research, Med. J. Aust. 210 (2019) S12–S16.
[46] M.L. Van Eck, X. Lu, S.J. Leemans, W.M. Van Der Aalst, Pm: a process mining project methodology, in: International Conference on Advanced Information Systems Engineering, Springer, 2015, pp. 297–313.
[47] S.Y. Arafat, S.K. Kar, M. Marthoenis, P. Sharma, E.H. Apu, R. Kabir, Psychological underpinning of panic buying during pandemic (covid-19), Psychiatry research.
[48] W. van der Aalst, Process Mining - Discovery, Conformance and Enhancement of Business Processes, Springer, 2011.
[49] A. Augusto, M. Dumas, M. La Rosa, M.A. Armas Cervantes, R. Conforti, M. Dumas, M. La Rosa, D. Reissner, Measuring fitness and precision of automatically discovered process models: A principled and scalable approach, Tech. rep., University of Melbourne (2019).
[50] R. Conforti, M.L. Rosa, A. ter Hofstede, Filtering out infrequent behavior from business process event logs, IEEE TKDE 29 (2).
[51] A. Augusto, A. Armas Cervantes, R. Conforti, M. Dumas, M. La Rosa, D. Reissner, Measuring fitness and precision of automatically discovered process models: A principled and scalable approach, Tech. rep., University of Melbourne (2019).
[52] R. Bose, R. Mans, W. van der Aalst, Wann improve process mining results?, in: 2013 IEEE (CIDM), IEEE, 2013, pp. 127–134.
[53] A. Rogge-Solti, R. Mans, W. van der Aalst, M. Weske, Improving documentation by repairing event logs, in: IFIP Working Conference on The Practice of Enterprise Modeling, Springer, 2013, pp. 129–144.
[54] J. Wang, S. Song, X. Lin, X. Zhu, J. Pei, Cleaning structured event logs: A graph repair approach, in: Proceedings of IEEE ICDE, IEEE, 2015, pp. 30–41.
[55] S. Song, Y. Cao, J. Wang, Cleaning timestamps with temporal constraints, VLDB Endowment 9 (101) (2016) 708–719.
[56] A. Augusto, M. Dumas, M. La Rosa, Meta-heuristic optimization for automated business process discovery, in: BPM, Springer, 2019.
[57] A. Augusto, A. Armas Cervantes, R. Conforti, M. Dumas, M. La Rosa, D. Reissner, Measuring fitness and precision of automatically discovered process models: A principled and scalable approach, Tech. rep., University of Melbourne (2019).
[58] R. Conforti, M.L. Rosa, A. ter Hofstede, Filtering out infrequent behavior from business process event logs, IEEE TKDE 29 (2).
[59] N. Tax, N. Sidorova, W.M.P. van der Aalst, Discovering more precise process models from event logs by filtering out chaotic activities, J. Intell. Inf. Syst. 52 (1) (2019) 107–139.
[60] M.F. Sani, S.J. van Zelst, W.M.P. van der Aalst, Improving process discovery results by filtering outliers using conditional behavioural probabilities, in: Proceedings of the Business Process Management Workshops, Springer, 2017, pp. 216–229.
[61] A. Baum, P.J. Kaboli, M.D. Schwartz, Reduced in-person and increased telehealth outpatient visits during the covid-19 pandemic, Ann. Intern. Med. 174 (1) (2021) 129–131.
[62] M. Mian, S. Sreedharan, S. Giles, Increased dispensing of prescription medications in australia early in the covid-19 pandemic, Med. J. Australia 214(9).
[63] P. Felli, A. Gianaola, M. Montani, A. Rivkin, S. Winkler, Cocomot: Conformance checking of multi-perspective processes via tant, in: International Conference on Business Process Management, Springer, 2021, pp. 217–234.