

# Stochastic Workflow Modeling in a Surgical Ward: Towards Simulating and Predicting Patient Flow

Christoffer O. Back<sup>1</sup>[0000-0001-7998-7167], Areti Manataki<sup>4</sup>[0000-0003-3698-8535],  
Angelos Papanastasiou<sup>3</sup>[0000-0002-5649-4551], and Ewen Harrison<sup>2</sup>[0000-0002-5018-3066]

<sup>1</sup> Department of Computer Science, University of Copenhagen, Denmark [back@di.ku.dk](mailto:back@di.ku.dk)

<sup>2</sup> Usher Institute, University of Edinburgh, Edinburgh, U.K. [mail@ewenharrison.com](mailto:mail@ewenharrison.com)

<sup>3</sup> Department of Mathematics and Statistics, University of Cyprus, Nicosia, Cyprus  
[apapan06@ucy.ac.cy](mailto:apapan06@ucy.ac.cy)

<sup>4</sup> School of Computer Science, University of St Andrews, St Andrews, U.K.  
[A.Manataki@st-andrews.ac.uk](mailto:A.Manataki@st-andrews.ac.uk)

**Abstract.** Intelligent systems play an increasingly central role in healthcare systems worldwide. Nonetheless, operational friction represents an obstacle to full utilization of scarce resources and improvement of service standards. In this paper we address the challenge of developing data-driven models of complex workflow systems - a prerequisite for harnessing intelligent technologies for workflow improvement. We present a proof-of-concept model parametrized using real-world data and constructed based on domain knowledge from the Royal Infirmary of Edinburgh, demonstrating how off-the-shelf process mining, machine learning and stochastic process modeling tools can be combined to build predictive models that capture complex control flow, constraints, policies and guidelines.

**Keywords:** Surgery · Surgical Workflow · Bayesian Network · Petri Nets · Simulation · Data Mining · Patient Flow · Process Mining

*Acknowledgements* We would like to thank Cameron Fairfield and Stephen Knight for their generous feedback regarding policies and on-the-ground practices at the Royal Infirmary of Edinburgh surgical ward. The final authenticated version of this paper is available online at [https://doi.org/10.1007/978-3-030-72379-8\\_28](https://doi.org/10.1007/978-3-030-72379-8_28)

## 1 Introduction

Surgical care is a key component of healthcare systems worldwide, saving and improving thousands of lives every day. Over 10 million operations are performed each year in England [34], including high-risk cases and patients that require immediate life, limb or organ-saving interventions. Surgical care is also very costly, with more than \$400 billion spent each year in the United States on operative procedures [1]. The number of people requiring surgery is rising every year, often leading to long waiting times that may put patients at risk.

Ensuring efficiency, timeliness and safety are crucial for providing high-quality service while controlling costs [26], [17]. While many processes surrounding surgery are

well structured, the dynamic nature of patient arrivals combined with the complexity of coordinating large numbers of specialized staff and facilities, means that delays and misalignments can have cascading effects leading to last-minute cancellations and under-utilization of expensive resources. There is, hence, an imperative need to improve surgical workflow. Some key questions here are: How can we improve overall surgical care performance in the most cost-effective way? How can we plan surgical care in a way that it is tailored to the individual patient?

There is a wealth of data being collected through hospital IT systems, which can be used towards answering these questions. This includes operating room management and usage data, electronic health records and surgery cancellation data. By adopting a process-based approach, one can make sense of such complex and big data and inform improvements in surgical care processes, including intelligent surgery planning, staff scheduling and workflow management.

This paper extends previous work [6] by presenting a preliminary investigation into stochastic workflow modeling and verification methods in surgical wards, with outset in a data set following patients from admission to discharge at the Royal Infirmary of Edinburgh in Scotland. With the aim of gaining a comprehensive understanding of surgical workflow, we use the data to investigate both *system-wide* surgical performance and *individual* patient flow. Results from these two types of modeling can be combined to enable personalised and efficient surgical scheduling.

In particular, we discuss how process mining methods can be used to gain insights regarding control-flow and temporal patterns in the surgical ward. Focusing on system-wide performance and recognizing the high level of uncertainty in the surgical department, we demonstrate how Stochastic Time Petri Nets can be used to effectively capture complex hospital policies and constraints. The choice of Stochastic Time Petri Nets allows for simulation of different scenarios, thus enabling what-if analysis. This is key for investigating different, and often competing, workflow improvement mechanisms. Focusing on individual patient flow, we propose the use of Bayesian Networks to predict patient-specific cycle times of individual surgical phases, from the time patients are sent for, through anesthesia and surgery, and until they leave recovery. Aside from their capacity to easily incorporate domain knowledge, Bayesian networks have the advantage that they can be queried in complex ways even with incomplete evidence, which is invaluable in the uncertain hospital environment. We present and compare three probabilistic models and we evaluate them w.r.t. to prediction accuracy. Crucially, we show that by incorporating a pre-processing step based on simple clustering of flows w.r.t. cycle times, we can improve the performance of our models noticeably.

The rest of the paper is structured as follows. In Section 2 we review existing literature. Our subsequent analysis of the data follows the classic data analytics workflow of *Describe*  $\rightarrow$  *Diagnose*  $\rightarrow$  *Predict*. In Section 3 we introduce the domain, the data set, and the data cleaning process. In Section 4, we present a descriptive analysis of the data set using process mining and standard statistical tools to identify control-flow and temporal patterns in the data. This informs the process of building system-wide simulation models and individual-patient predictive models, which we describe and evaluate in Section 5. In Section 6 we discuss our results and in Section 7 we conclude and discuss directions for future work.

## 2 Related Work

The modeling of surgical workflows has received a significant amount of attention by researchers, motivated by the prerogative to improve efficiency and resource utilisation while ensuring adherence to service standards.

Of particular interest for the present case study is the National Theatres Project in Scotland which outlines several areas for improvement that might be addressed by workflow optimization. This includes “appropriately increasing patient throughput, thereby using resources more productively and efficiently” by reducing unutilized (operating room) hours; reducing over/under-runs, late-starts, cancellations and delayed discharges; and avoiding unnecessary out-of-hours and nighttime procedures [32].

Previous research on modeling surgical ward processes varies greatly in terms of scope: from very fine-grained models of individual procedures to high-level models of treatment pathways well beyond the context of the surgical ward itself (e.g. from visit to a GP to follow-up evaluation and treatment). In their literature review on the topic [26] Laylis and Jannin identify a range of granularity. At the finest level are low-level physical movements such as tool usage patterns based on sensor data [4], phase detection [37], automatic identification of hand motions from video in [27] and [9]. Several investigations have been made into the modelling of Cholecystectomies, a highly standardized procedure [11,13,12,31].

In [36] the authors go beyond the modeling of the surgical procedure to include anesthesia and early recovery within the operating theatre, while [20] considers the patient flow from admission to recovery. Activities downstream from surgery, namely recovery in ICU wards can present a key bottleneck, as addressed in [5]. Extending the patient pathway further, follow-up post-surgery is incorporated in [28,21]. In general, however, most research appears to have focused primarily on either very low-level procedure or high-level treatment pathways. The patient flows we consider fall in between these levels of granularity.

The use of Bayesian networks to model stays in an emergency department is evaluated in [3]. In contrast to our approach, the view of patient flows is at a higher level of abstraction, and the main focus is the comparison of structure learning algorithms.

Modeling the duration of surgical procedures was investigated in [38,24] and we are able to report findings in line with these regarding the log-normal distribution of surgical times. Surgical duration was incorporated into sequencing and scheduling strategies in [17]. Stochastic balancing of bed capacity based on fluctuating demand patterns was explored in [16] and length of stay patterns in [5] while resource allocation and patient admission was addressed in [22].

More broadly, the problem of ensuring that systems fulfill a given set of specifications has been widely addressed in model checking and process mining research. The systems under consideration can range from electronic circuits and communication protocols [33,29], to business processes [23,15] and even biological systems [18].

We can verify whether a system satisfies a property or specification by means of state-space exploration [40] or rewrite rules [8], but often realistic models reach a level of complexity that precludes closed-form verification. This leaves simulation, essentially a random sampling of the model’s state space, as a next best tool for verification and what-if analysis.

Incorporating desired constraints into a system model is straightforward [7], and allows for correct-by-construction plan generation, but also leads to state-space explosion. Advancements in seemingly unrelated areas such as robot motion [19] provide evidence that this approach to intelligent planning is feasible, even in complex domains.

### 3 Domain and Data Preparation

We were granted access to workflow data recorded at The Royal Infirmary of Edinburgh in connection with cases taking place from 2010 until 2018. The infirmary is Scotland’s largest, with 900 beds and a 24-hour accident and emergency department. The data at our disposal was recorded by the Operating Room Scheduling Office System (ORSOS), which is one component in the institution’s overall IT-infrastructure.

Over 1700 types of procedures are recorded in the data set, about half of which are classified as emergency cases. Each treatment procedure is given a unique case ID, meaning that the same patient may be associated with multiple case IDs, even during the same stay for inpatients. Following patients’ broader treatment patterns would be possible using this dataset, but lies beyond the scope and focus of this paper.

Data is entered manually by surgical support personnel, with the system requiring the entry of timestamps for each event in the patient flow. Figure 1 illustrates the proscribed sequence of events, along with an aggregation of activities into logical phases (pre-op, anesthesia, surgery, recovery).

The system attempts to enforce a strict linear ordering of events, though this can be overridden by personnel. If a timestamp is entered out of sequence, a warning is given, but can be entered upon confirmation. Staff are then sent a summary of anomalous cases for review at the end of the week.

*Data Schema* In addition to timestamps for the 11 proscribed activities in a patient flow, 34 other attributes are recorded. Attributes of note include two different procedure coding schemes, case type (emergency/scheduled, day-case/inpatient), NCEPOD urgency classification<sup>5</sup>, and the ASA patient condition rating.

Some staffing details are also included, such as main and supervising surgeon and anesthetist, as well as the consultant assigned to the case. The source of admission (emergency room, etc.), as well as intended and actual destination following surgery (ICU, etc.) and crucially, the operating room number, are also included. Further details include the diabetic status of the patient, types of anesthetics administered, whether antibiotics were administered, and whether pre-session briefings and surgical pauses were held.

*Cleaning and Preparation* A number of clearly anomalous entries are present in the dataset, comprising roughly 10% of the 38,728 entries. Due to the relatively small percentage of anomalies and the reasonably large dataset, we followed a precautionary principle and simply removed entire cases containing anomalous entries prior to further analysis and modeling. Table 1 provides an overview.

<sup>5</sup> NCEPOD Classification of Intervention [30].

ANOMALY	COUNT	% OF TOTAL
Duplicate entries	58	0.15
Missing values	31	0.08
Dates out-of-range	475	1.23
Zero timestamps	3089	7.98
Bad ordering	443	1.44
Total	4096	10.58

Table 1: Anomalous cases removed prior to analysis. Originally published in [6].

Duplicate entries may have been due to an attempt to correct a data entry error, but we are unable to determine which entry is reliable. The column *anaesthetic start time* was the only timestamp column to contain NA values. A larger number of cases have clearly anomalous values in the case date column, e.g. dates much too far in the past (year 1800) or future (year 3206).

## 4 Analysis

### 4.1 Control Flow Patterns

Based on input from domain experts, we were aware of the *de jure* workflow, which follows a simple linear flow of events as illustrated in Figure 1. In addition to the anomalies discussed in Section 3, process mining techniques helped reveal further control-flow deviations, guiding the data cleaning process. In particular, we found dotted charts, directly-follows graphs and mined Petri Nets to be particularly informative.

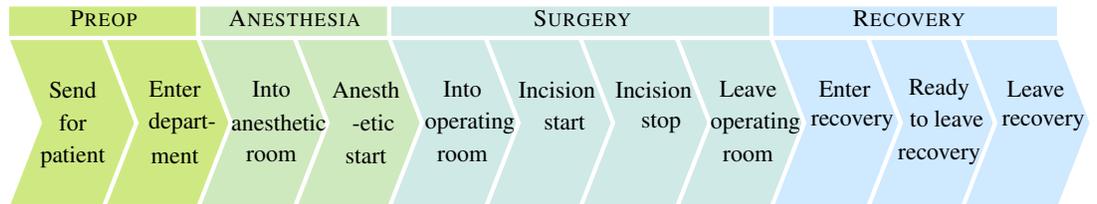


Fig. 1: The *de jure* sequence of events recorded by the ORSOS system, representing the intended patient flow. Activities are linearly ordered, but can occur “simultaneously”. That is, some activities (such as *Leave Operating Room*) can have the same timestamp as the “succeeding” activity (*Enter Recovery*), but should not occur after it. Originally published in [6].

*Dotted Chart* One simple yet powerful tool for getting a quick, preliminary overview of process-related data is the dotted chart, which simply charts events w.r.t. case-id across

time such that dots falling along a horizontal line represent events belonging to the same process instance (i.e. case).

Using the dotted chart in Figure 2 we immediately identified a substantial gap in the dataset. Furthermore, we can see that some cases have events occurring many months, even years apart - almost certainly evidence of anomalous entries.

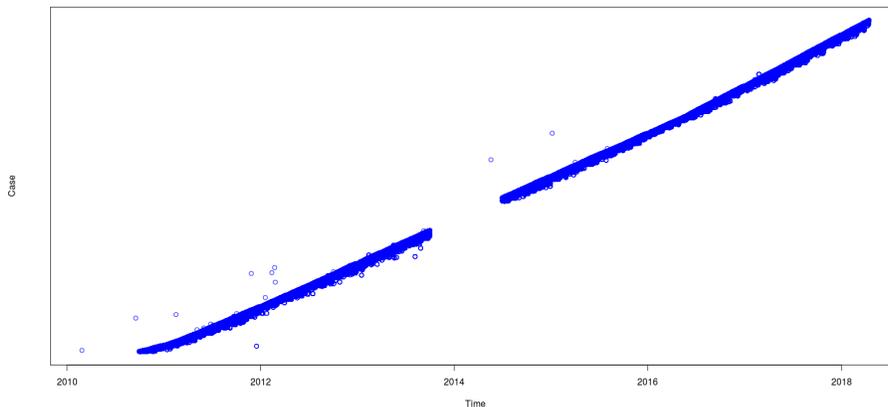


Fig. 2: A dotted chart showing all events in our dataset, arranged according to case id and timestamp. Originally published in [6].

*Directly-Follows Graph* Another simple visualization tool, directly-follows graphs consist of nodes and directed edges, where nodes represent the events in the log, and an edge exists between two nodes if there is at least one log trace where the source event is followed by the target event. Figure 3 shows the directly-follows graph obtained for our dataset, which includes node and edge frequencies. On one hand, the event frequencies on the graph confirmed that all events were included in each trace, in accordance with the *de jure* workflow. On the other hand, the graph indicated that nearly all possible pairwise event orderings occurred at least once in the data. This is inconsistent with the *de jure* workflow, and it includes several implausible event orderings. For example, there were a remarkable 154 traces where the last event in the *de jure* workflow, namely *leave\_recovery\_time*, occurs before the proscribed first event, namely *sent\_for\_time*.

*Alpha Miner* For a more nuanced view of the control-flow evidenced by the event log, proper process mining algorithms can be used. The Alpha ( $\alpha$ ) miner was one of the first process mining algorithms developed, and though it has limitations regarding the variety of control constructs it is able to identify, for our purposes it provided interesting insights into course of events as evidenced by the data.

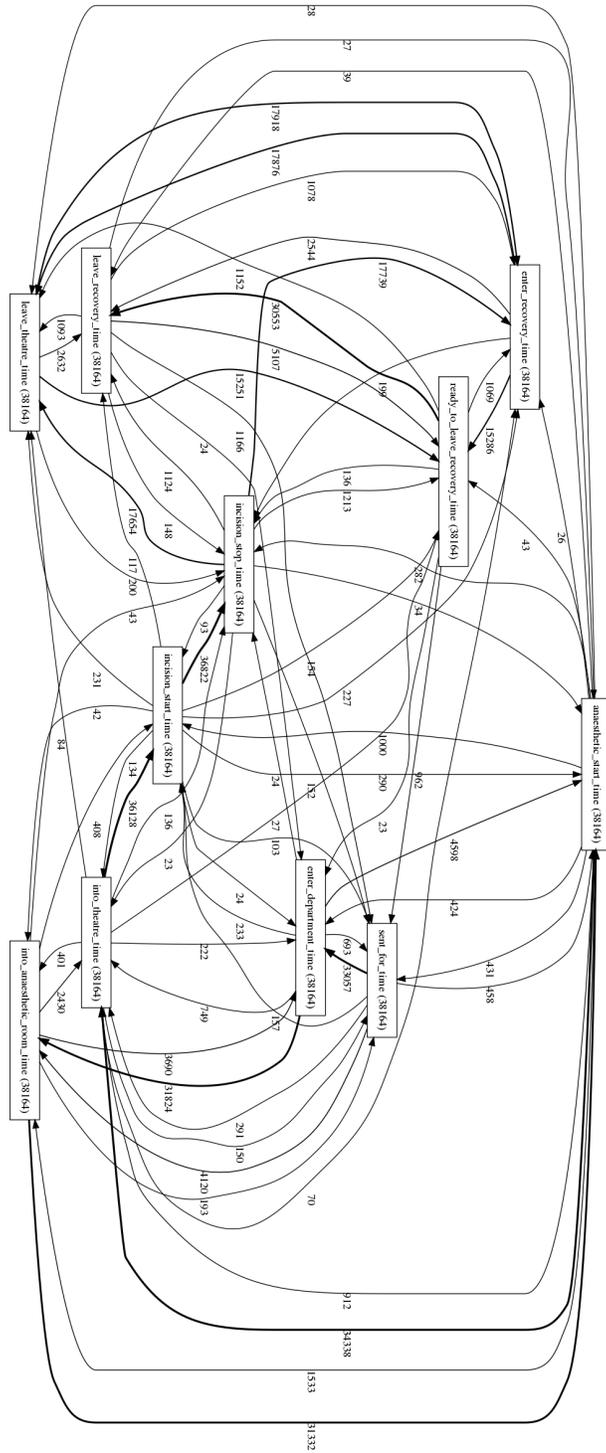


Fig. 3: Directly-follows graph indicating that nearly all possible pairwise event orderings occurred at least once in the data.

Figure 4 shows the result of running the SIMPLE version of the Alpha miner[2] from the pm4py package [10] on the top 20 sequence variants in the log. Mining on the entire log produces a flower model - a model which permits any behavior, in line with observation from the directly-follows graph in Figure 3.

According to this model, several remarkable control patterns seem to be evidenced by the most frequently occurring sequence variants. For example, according to Figure 4, *anesthetic\_start* is not a precondition for *incision\_start*. This observation led us to inquire with experts at the infirmary and to more closely investigate these cases in the dataset. Apparently, it was not uncommon for these two events to have exactly the same timestamp: a reflection, for example, of cases in which a surgeon administers a local anesthetic immediately prior to a minor surgery.

This observation gives rise to a further insight: nearly all process mining algorithms have a strong assumption of temporal monotonicity, i.e. events are strictly linearly ordered such that no two events share *exactly* the same timestamp. With the coarse level of temporal accuracy (1-5 minutes) in our dataset, exactly co-occurring events were common. In this sense, most process mining algorithms are unable to account for *true* concurrency in data.

## 4.2 Temporal Patterns

Beyond identifying anomalies in the data, there were few interesting control-flow patterns to identify at the level of patient flows, since they do in fact follow a (non-monotonic) linear process.

This left temporal patterns as the next obvious aspect to investigate, especially

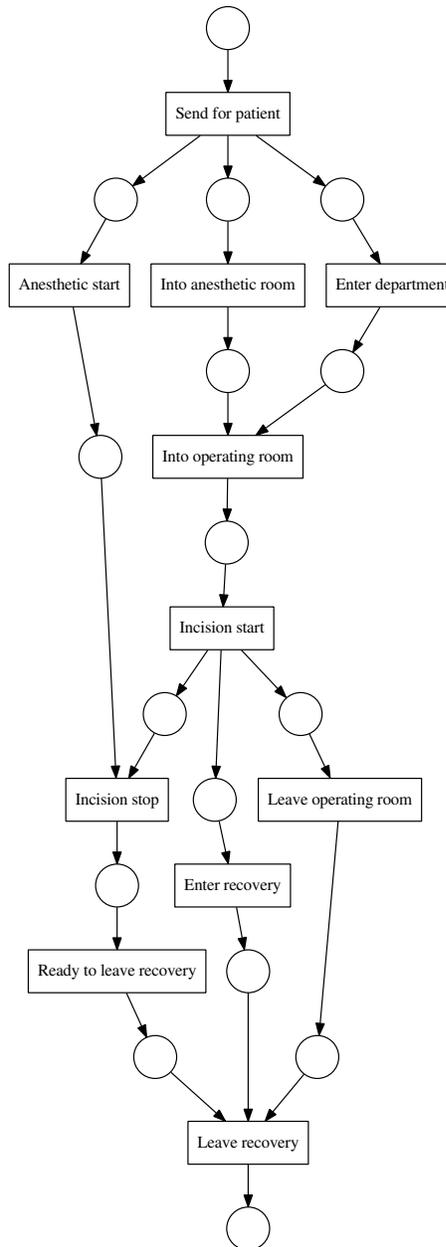


Fig. 4: Petri net generated by the Alpha miner on the top 20 trace variants. Originally published in [6].

since both resource usage and service guidelines are largely temporally focused (e.g. target time to theatre, anaesthetist availability).

*Event Aggregation* In part due to large numbers of zero-duration cycle-times due to the phenomenon of co-occurring events, but also based on conversations with experts, we decided to group individual events into the four phases illustrated in Figure 1. Aside from clearly representing logically meaningful phases, it was also clear that this aggregation smoothed out cycle-time distributions.

On one hand, this process of aggregation arguably *removes* valuable information that could inform our model, on the other, it constitutes a form of dimensionality reduction which helps control the complexity of our model and ultimately improves performance in the end.

*Marginal Distributions* The simplest temporal pattern at the level of patient flows is the marginal distribution of cycle times across all patients regardless of procedure, condition, urgency, etc. Fitting a probability distribution to the empirical distributions of cycle time also constitutes the simplest possible predictive model, i.e. the maximum likelihood prediction based on the best-fit distribution.

In Table 2 we show the goodness-of-fit statistics (Kolmogorov-Smirnov criterion) for 7 types of distributions for both the individual events and events aggregated into phases. Those distributions best fit to the phase data are visually depicted.

In building the stochastic Petri net described in Section 5, we used these marginal distributions to parametrize transitions representing these phases. However, our modeling tool restricted the families of probability distributions to Exponential, Erlang and polynomials of exponentials. We illustrate our approximations to the best-fit distributions in Table 2 and give the exact parametrizations in Table 5.

*Mutual Information* To get an impression of which attributes might be informative independent variables in conditional distributions of cycle times, we calculated estimates of pairwise mutual information. Having an eye to identifying variables for inclusion in the Bayesian networks described in Section 5.2, we were interested in mutual information between all attributes.

As a measure of the expected decrease in uncertainty regarding the outcome of variable  $X$  upon learning the outcome of  $Y$ , mutual information is akin to standard correlation metrics, but well suited to hybrid (discrete/continuous) attributes and makes no assumption regarding normality or linearity.

Recalling the definition of the Shannon entropy of a random variable  $X$  as its expected information content, denoted  $H[X]$ , we can write mutual information directly as the decrease in entropy of  $X$  upon learning the outcome of  $Y$ . Formally,  $I(X; Y) = H[X] - H[X|Y]$ . Two completely independent variables will have mutual information of  $H[X] - H[X] = 0$ , while for two perfectly correlated variables it collapses to the entropy of the dependent variable  $H[X] - 0 = H[X]$ .

However, as an expected value (averaged over the sample space), it can hide that specific outcomes for a variable can have a high *pointwise mutual information* - which could be harnessed by a predictive model - yet disappear amongst many uninformative

	Gaussian	Cauchy	Logistic	Log-Normal	Gamma	Weibull	Exponential	
EVENT	GOODNESS-OF-FIT (KS)							PLOT (Best fit for aggregate)
Send for patient	0.147	0.113	0.139	0.169	0.126	0.104	0.267	
Enter department	0.161	0.147	0.197	0.205	0.171	0.157	0.184	
<b>Pre-op</b>	0.094	0.087	0.123	0.127	0.09	<b>0.062</b>	0.24	
Into anesthetic	0.226	0.166	0.168	0.153	0.133	0.15	0.19	
Anesthetic start	0.146	0.098	0.134	0.171	0.112	0.096	0.189	
<b>Anesthetic</b>	0.124	<b>0.077</b>	0.106	0.188	0.132	0.106	0.244	
Into theatre	0.16	0.094	0.143	0.111	0.093	0.114	0.298	
Incision start	0.164	0.122	0.144	0.061	0.06	0.07	0.132	
Incision stop	0.187	0.145	0.168	0.111	0.144	0.128	0.25	
<b>Surgery</b>	0.16	0.11	0.134	<b>0.036</b>	0.071	0.087	0.193	
Enter recovery	0.083	0.079	0.126	0.243	0.174	0.139	0.198	
Ready to leave	0.285	0.277	0.266	0.184	0.144	0.144	0.22	
<b>Recovery</b>	0.099	<b>0.083</b>	0.127	0.244	0.17	0.136	0.19	

Table 2: *Red*: Best fits for marginal distributions of cycle times, goodness-of-fit statistic used is the Kolmogorov-Smirnov criterion. *Blue*: distributions used for modelling, in which our tool restricted the choice of distributions to Erlang (pre-op) and polynomials of exponentials (remaining). Originally published in [6].

outcomes. For this reason we also manually explored conditional distributions for cycle times.

*Conditional Cycle Time Distributions* Our investigation around the most informative features in the data set continued by exploring the conditional distributions of cycle times for the individual values attributes. By visualizing conditional distributions on the same plot, one gets a quick impression of whether an attribute is informative in this respect, or not. Even though this is a somewhat time-consuming, brute-force approach, exploring the data in this way turned out to be quite informative. This analysis played an important role for us in choosing which variables to include in the models we present in Section 5.2. Examples of some of the most informative attributes are presented in Figure 5. For instance, one can see that the conditional cycle time distributions for surgery differ considerably based on ASA status, i.e. for normal healthy patients (ASA status 1), for patients with severe systemic disease that is a constant threat to life (ASA status 4) and for patients with non-assessed ASA status. The anesthetic cycle time distributions

conditioned on source of admission also differ considerably, with patient cases coming from the High Dependency Unit spending longer on average in Anesthesia, compared to those coming from the Admissions Unit or some other source.

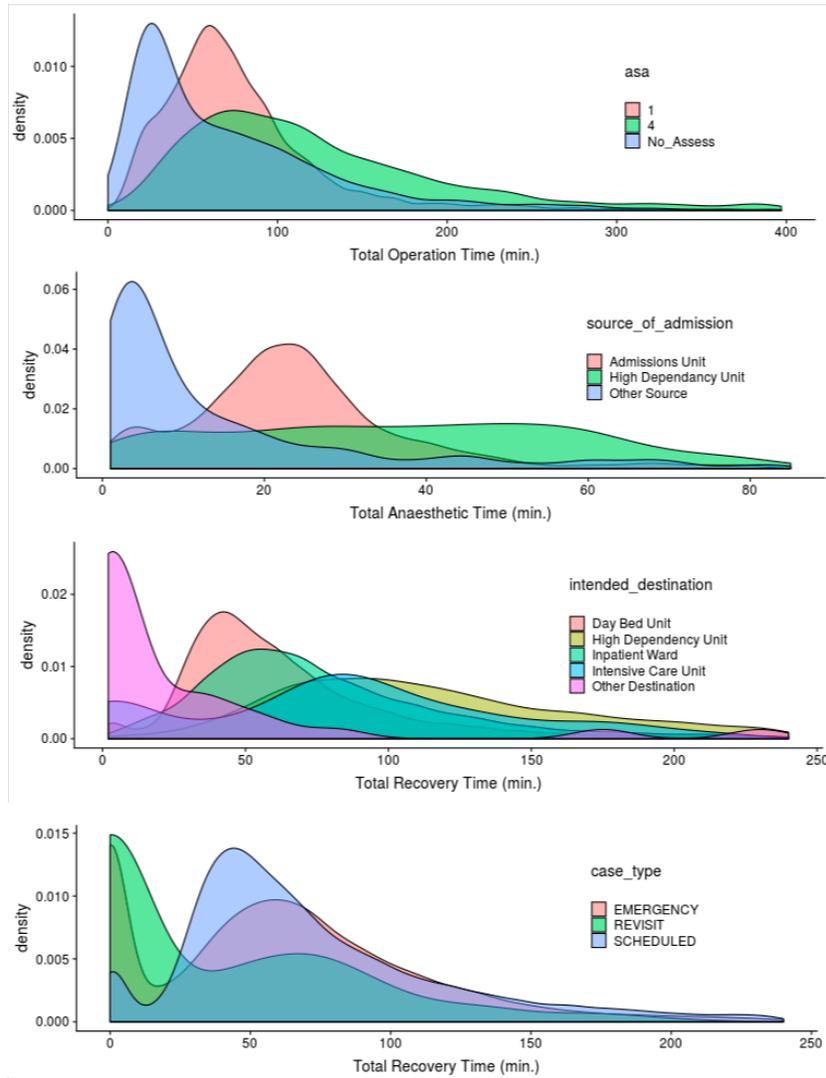


Fig. 5: Examples of conditional cycle time distributions. Top: conditioned on ASA status. Second from top: Source of Admission. Third from top: Intended Destination. Bottom: Case Type.

*Principal Component Analysis & Patient Clusters* We explored the presence of groupings of patients in regards to duration by a combination of visual analysis, data transformation and clustering.

Judging by the original durations of the 4 phases of a patient’s flow, there do not appear to be clear groupings of patients (Figure 6a). However, after applying principal components analysis (PCA) and plotting the data w.r.t. the four principal components, clear groupings become apparent (Figure 6b). Since PCA assumes normally distributed data, and since most durations more closely follow a log-normal distribution, the data was log-transformed prior to PCA transformation.

Afterwards, *k*-means clustering was used to discover grouping of patients. This derived attribute was added to the dataset and our predictive models, noticeably improving performance. It should be noted that we retained all 4 principal components, and thus employed PCA solely as a *transformation* rather than *dimensionality reduction* technique, as is common. This was due to the observation that removing those principal with lowest eigenvalues did not improve performance. This is not unexpected, considering the small number of dimensions.

### 4.3 Arrival Rates

Many of the aspects of patient flows we have considered so far concern patterns at the level of the individual patient. In order to model system-level dynamics it is crucial to consider how the system is affected by multiple processes competing for shared resources.

One key component in this analysis concerns the arrival of patients, in particular unplanned arrivals requiring immediate treatment, since this will affect and potentially interfere other patient flows. A clearly defined policy exists for the prioritisation of cases based on severity which can lead to cancellation of procedures.

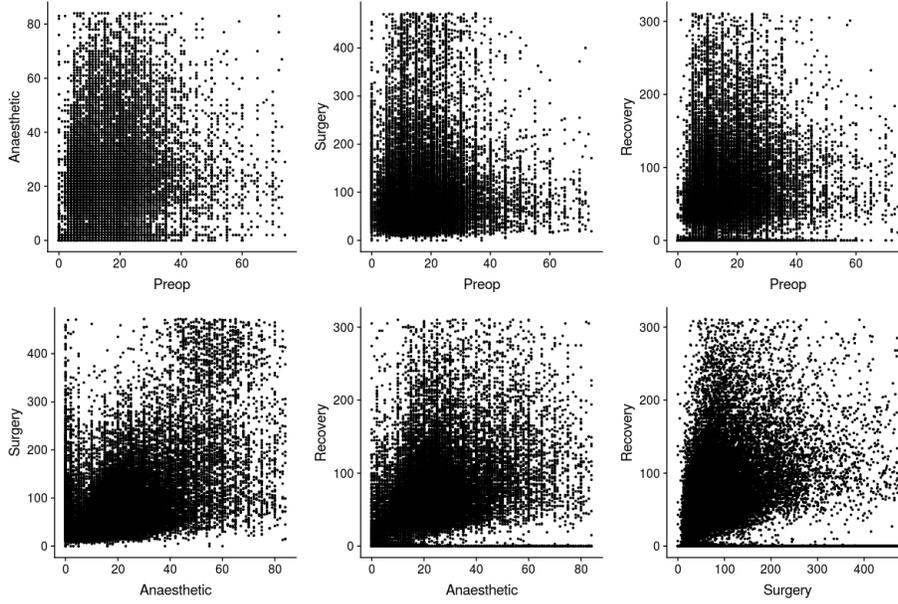
A common assumption in performance modelling and queuing theory is that the number cases arriving for service within some interval follows a Poisson distribution. We found those cases arriving via the emergency room (unscheduled) did in fact follow a Poisson distribution remarkably well (Figure 7) whereas scheduled cases did not. The latter is not so surprising since the arrival of scheduled cases is necessarily a non-random process and is adjusted to balance the arrival of emergency cases.

The close fit of daily emergency arrivals to a Poisson distribution allows us to accurately model the remaining stochastic transition in our model (the other representing marginal cycle times) using the closely related exponential distribution, which captures the corresponding distribution of inter-arrival times.

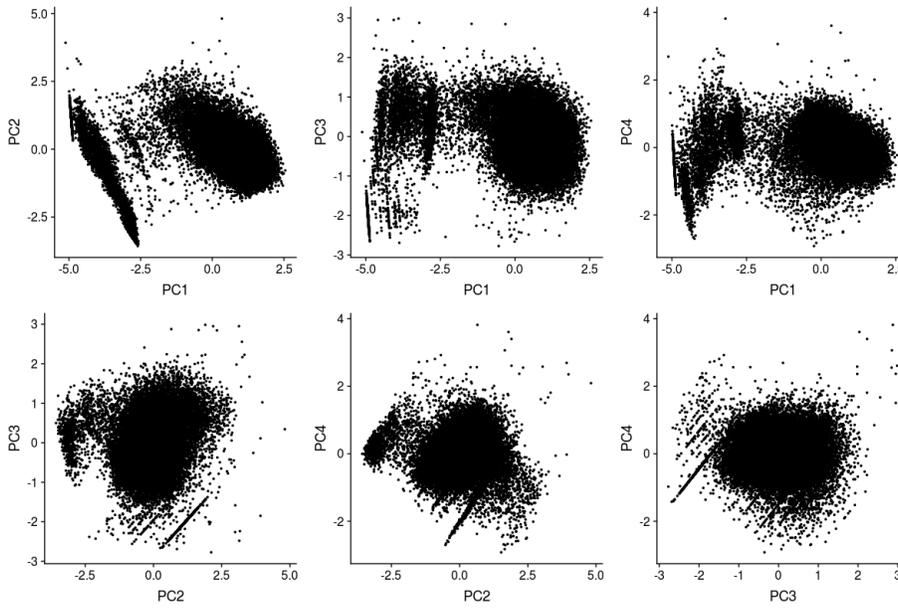
## 5 Modeling

### 5.1 Stochastic Time Petri Nets for System-Wide Simulation

To model the surgical workflow, we used Stochastic Time Petri Nets (STPN) which are essentially Petri nets in which the notion of time and uncertainty is incorporated by adding either a deterministic or a probabilistic delay for the firing of transitions. Specifically, transitions can be either immediate, deterministic or stochastic (Table 3). The



(a) Raw durations of 4 phases plotted against each other



(b) Durations w.r.t. principal components. Data was log-transformed prior to PCA transformation.

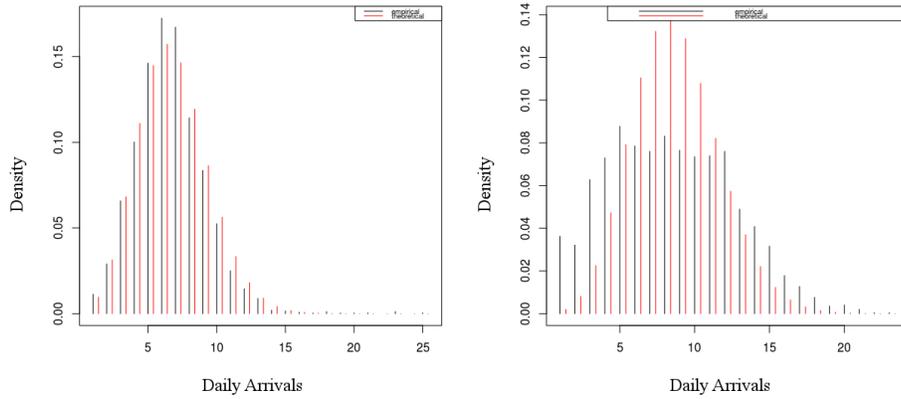


Fig. 7: Daily arrivals (black) along with best-fit Poisson distribution (red). *Left*: emergency cases. *Right*: scheduled cases.

model was implemented in ORIS API; a software tool for the modelling and evaluation of stochastic processes [14]. By using the functions provided by the ORIS tool we were able to evaluate the performance of the system by observing how different metrics change when we alternate some of its aspects.

TRANSITION	REPRESENTATION	DESCRIPTION
Immediate	? condition 	Fires immediately if enabled, conflicts between competing enabled immediate transitions are resolved using priority ranking
Deterministic	5 █	Fires after a fixed amount of time upon becoming enabled
Stochastic	$\lambda e^{-\lambda x}$ █	Fires after a delay sampled from a probability distribution upon becoming enabled

Table 3: Overview of transition types in Stochastic Time Petri nets.

*Surgical Ward Workflow Description* The scenario that is considered for this study is the following: Emergency patients arrive in the hospital at a certain rate to receive treatment throughout the entire 24 hour period (transition *arrival* in Figure 8b) while elective ones are only allowed to arrive at the hospital during the working hours (uniform transition *arrival* in Figure 8a with enabling function `?working_hours=1`). Emergency arrivals go through a checking procedure for the determination of the severity of their condition. For the current model we assume that about half of the cases do not require

immediate intervention (uniform transitions *emergency\_status* and *scheduled\_status*). If it is decided that the operation must be performed immediately, the patient moves to the Preop room in order to get prepared for the operation (place *preop\_room*).

According to the NCEPOD urgency classification [30], target time to theater varies depending on the case. For the purposes of this study, only one type of emergency is considered, and the expected time to theater was set to 30 minutes. On the other hand, scheduled cases can be cancelled up to the time they are about to be placed under anaesthesia if there are no available resources (e.g. beds, surgeons, theaters) to continue the process (transitions *cancel\_1*, *cancel\_2*). This is not the case for emergency patients, however, who can move from phase to phase (pre-op, anesthesia, surgery, recovery) if the resources for the next part of the process are available. In this paper, we assume that a surgeon and an anesthetist is required for the surgery. Furthermore, the anesthetist is also required during anesthesia and recovery. Note that priority is given to emergency patients over scheduled ones when a decision has to be made regarding the entry to an operating theater or an anesthetic room. The duration of each phase (pre-op, anesthesia, surgery, recovery) is modelled using stochastic transitions with random firing rates following distributions that match our findings in Section 4 (Table 2). The properties of these transitions is shown in Table 5.

To account for cancellations and delays, two places were added in the STPN, namely *cancelled\_cases\_24h* and *emergencies\_wait\_cases*. In the former a token is added every time a scheduled case gets cancelled while in the latter a token is added whenever an emergency case is waiting for more than 30 minutes. Both places are reset to zero at the start of a working day. Prioritization and availability checking were incorporated in the model by setting the proper enabling functions and marking updates to transitions. For instance, the enabling function of the transition *enter\_recovery* was set to *bed\_available > 0*. This property of the model prevents emergency patients to enter the recovery phase if there are no available beds. Table 4 illustrates some examples of how our STPN captures some other policies and guidelines.

Additional simplifying assumptions made for this first iteration of modeling include: that anesthetic rooms and operating theatres are completely independent, a constant number of resources are available within and outside working hours and only one type of recovery room is present whereas in reality different sections are present, such as ICU and high-dependency unit.

Figure 8c shows the basic outline of our model. Patient flows were modeled as individual workflow nets in order to capture constraint violations *for individual patients*. These workflow nets were then programmatically duplicated during simulations. Due to space limitations we are unable to elaborate all details of the model, particularly enabling/update functions and firing priorities, but the full model can be found online <sup>6</sup>.

*Simulation* Using our model, we investigated different resource capacity scenarios (Table 6), with a focus on the ability of the system to fulfill two quality-of-service indicators:

- Number of scheduled cases cancelled in 24 timeframe

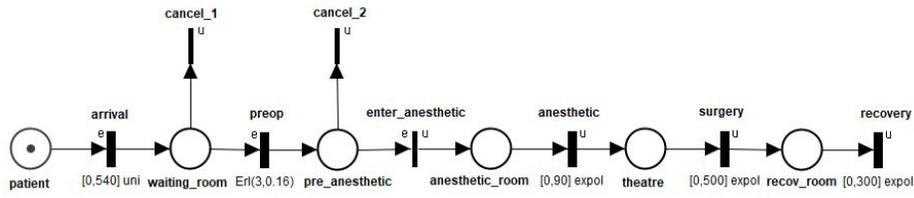
<sup>6</sup> <http://www.github.com/apapan08>

POLICY/GUIDELINE	PETRI NET FRAGMENT
Normal working hours are from 8-18	
Non-emergency procedures should be handled within normal working hours	
Scheduled cases can be cancelled all the way up to entering anesthetic room if all theatres are occupied by higher priority cases	<p>PRIORITY: enter anesthetic &gt; cancel</p>

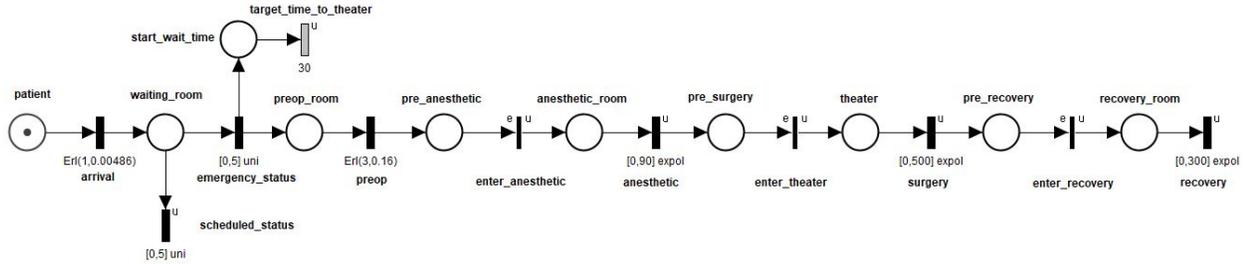
Table 4: Simplified examples of some of the policies, guidelines and constraints for patient flows captured by our model, along with the fragment of the Petri net which captures this.

TRANSITION	DISTRIBUTION	PARAMETERS
Emergency arrivals	Exponential	$\lambda = 0.00487$
Preop	Erlang	$k = 3, \lambda = 0.16$
Anesthetic	Polynomial of Exponentials	$1.5x^2e^{-0.11x} - 10xe^{-0.11x} + 30e^{-0.11x}$
Surgery	Polynomial of Exponentials	$x^2e^{-0.0385x} + xe^{-0.0085x}$
Recovery	Polynomial of Exponentials	$x^2e^{-0.05x} - 10xe^{-0.1x} + 150e^{-0.1x}$

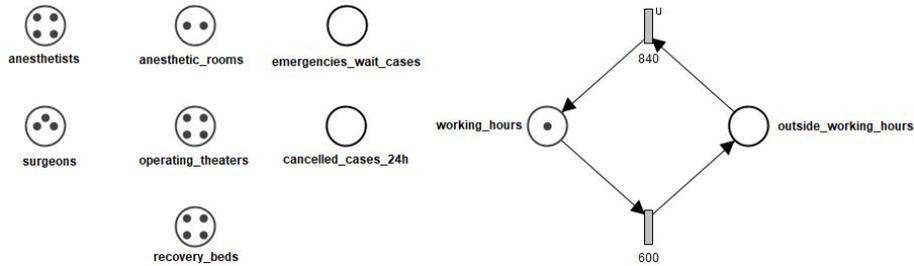
Table 5: Fitted parameters for stochastic transitions. See Table 2 for a visualization.



(a) Template for scheduled patient flow.



(b) Template for emergency patient flow.



(c) Shared resources.

Fig. 8: Stochastic Time Petri Net used to model core aspects of patients flows.

- Target time to theatre < 30 minutes for emergency cases

These are properties that are straightforward to formalize and evaluate. In fact, any properties that can be formalized in an appropriate temporal logic such as LTL or MITL<sup>7</sup>, which read similarly to natural language guidelines, can be evaluated. In addition to reporting the expected values for these QoS criteria, we included the expected resource availability over time. This helps us to identify when, and for which resources, potential bottlenecks arise. In a more sophisticated model, we would likely see more complex patterns resulting from interacting processes/resources.

<sup>7</sup> Linear Temporal Logic, Metric Interval Temporal Logic.

	EXCESS CAPACITY	SUFFICIENT CAPACITY	INSUFFICIENT CAPACITY
scheduled cases/day	10	10	10
anesthetists	10	8	4
surgeons	8	5	3
anesthetic rooms	8	4	2
operating theatres	10	8	4
recovery beds	12	6	4

Table 6: Resource capacity scenarios explored in simulations.

We report results in terms of expected value<sup>8</sup> across 100 simulation runs of each scenario. That is, the number of cancelled cases or available operating theatres at a given time, averaged over simulation runs. With this simple model, the state of the system follows a consistent periodic fluctuation according to the working hours (Table 7). Realistically, however, greater fluctuations would be likely due to uneven patient inflows and staffing availability patterns over time.

## 5.2 Bayesian Nets for Individualised Prediction

In modeling cycle times in patient flows, our model in Table 2 considers only marginal distributions, i.e. cycle time estimates are identical for all patients. However, as we illustrated in Section 4.2, there clearly exist categories of patients with significant variations in cycle times.

Bayesian Nets are probabilistic graphical models that capture the structure of complex probability distributions. By exploiting conditional independence relations between variables, inference algorithms allow us to query the belief network in a flexible manner, even with only partial information [25]. In the present context, Bayesian nets allow us to take into account multiple attributes of a patient along with the partial completion of their treatment in order to make significantly more accurate and nuanced predictions regarding cycle time and other aspects, such as destination.

We present the results of two Bayesian networks in predicting cycle time, leaving as an important avenue of future research the integration of these predictive models into a more sophisticated process model which accounts individual patient attributes. Figure 9 illustrates that Bayesian networks can be integrated with Stochastic Petri net models to more accurately model transition distributions specifically.

The models were built and trained using algorithms implemented in the `bnlearn` package for the programming language R [35]. We used our own implementation of cross-validation, in part to avoid data snooping, and added simple smoothing procedures to account for undersampling.

*Feature Selection* We evaluated two networks: a 10-variable model and a 22-variable extension of the first. The decision of which attributes to include was based in part on our exploration of conditional distributions in Section 4.2 and pairwise mutual information described in 4.2.

<sup>8</sup>  $E[X] = \sum x p(X = x)$

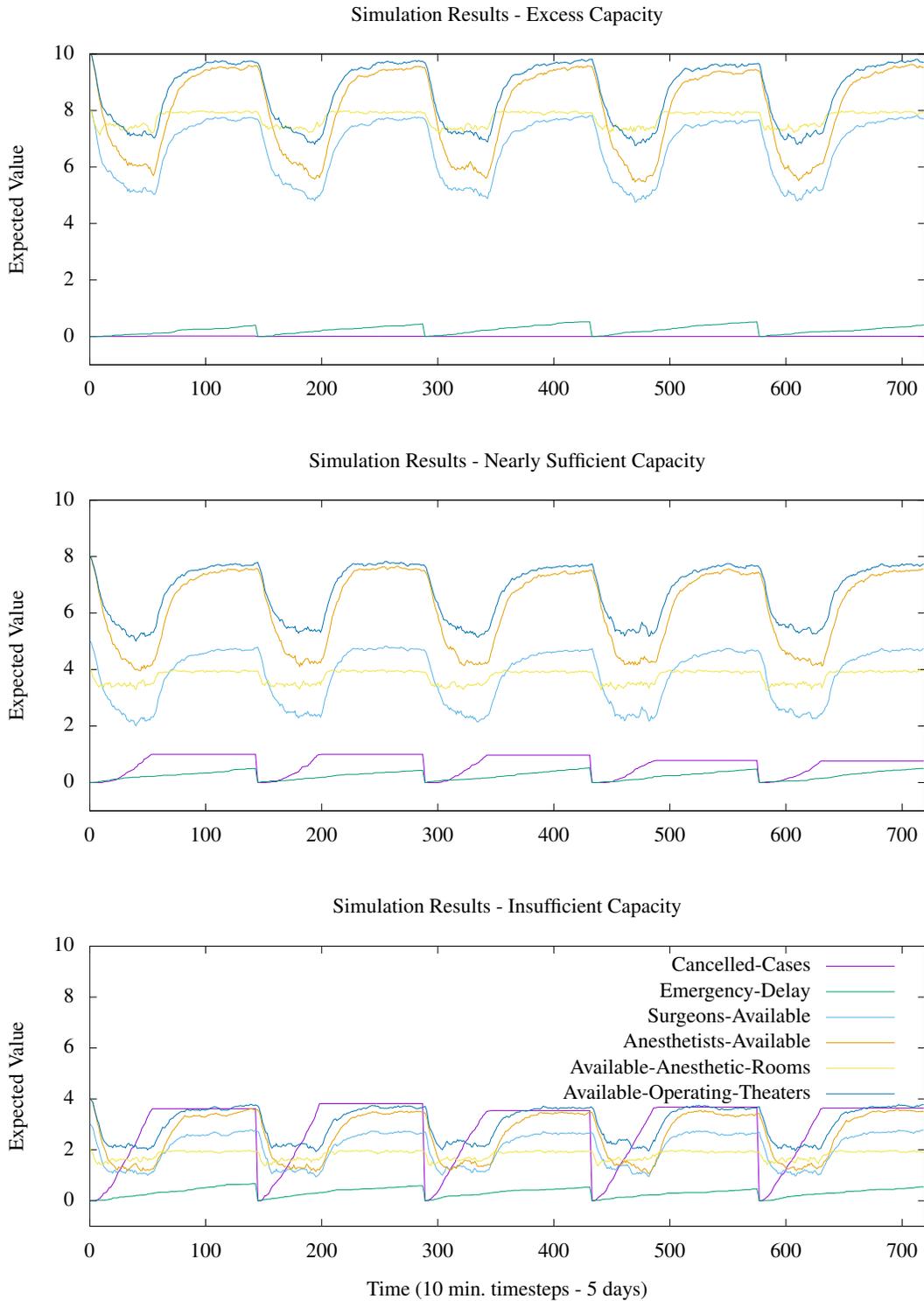


Table 7: Simulation results for scenarios with (*top*) excess capacity, (*middle*) sufficient capacity, and (*bottom*) insufficient capacity averaged over 100 runs. Cancelled-Cases and Emergency-Delay represent failures to meet quality of service guidelines when the expected value exceeds 0. In other words: these are the system properties we are interested in verifying.

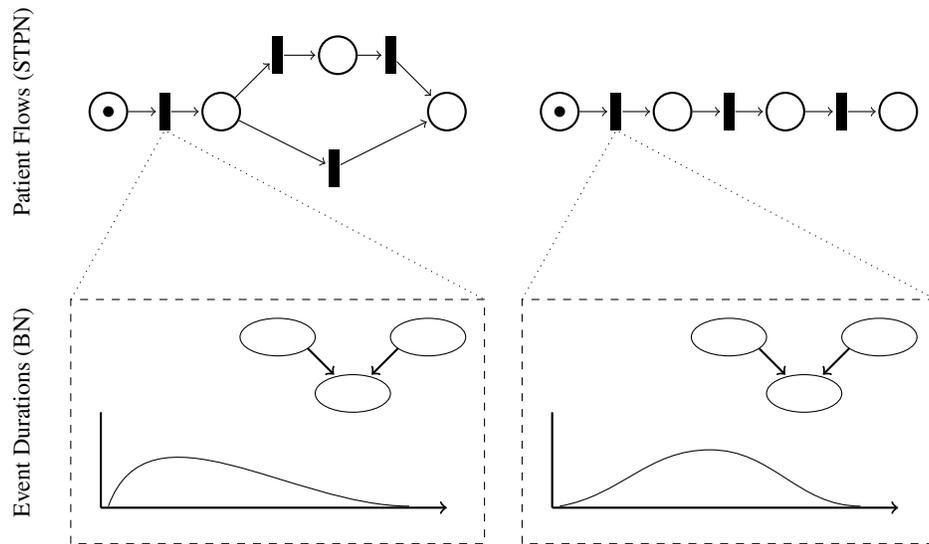


Fig. 9: Petri nets capture control-flow of a process, while Bayesian networks allow nuanced modelling of transition distributions based on case attributes. Integrating these two modelling perspectives is an important next step in developing data-driven, patient centric workflow models.

One feature of note is the *Cluster* node in both 10 and 22 variable models. This represents the patients groups identified in 4.2. Using simple  $k$ -means clustering, we experimented with identifying 5, 10, 15, and 20 patient clusters which proved moderately helpful in improving performance - in particular for predicting anesthetic cycle times using the 22-variable model.

*Structure Learning* There are two methods for constructing the graph structure of a Bayesian net: manually, based on expert knowledge; and automatically using structure learning algorithms. After several attempts at building nets manually, we found that automatically generated nets outperformed, despite sometimes finding odd connections between variables.

We employed the score-based structure learning algorithms hill-climbing and TABU structure-learning algorithms, using scoring functions Akaike and Bayesian Information Criterion (AIC/BIC). This choice was due to their computational tractability and suitability to the our hybrid dataset (continuous and discrete attributes). The 22 variable graph can be seen in Figure 10.

*Evaluation* Models were evaluated based on prediction of cycle times for the 4 phases of the surgical patient flow using 10-fold cross validation. We present mean *absolute* error in Figure 11 for comprehensibility, but note that mean squared error results closely

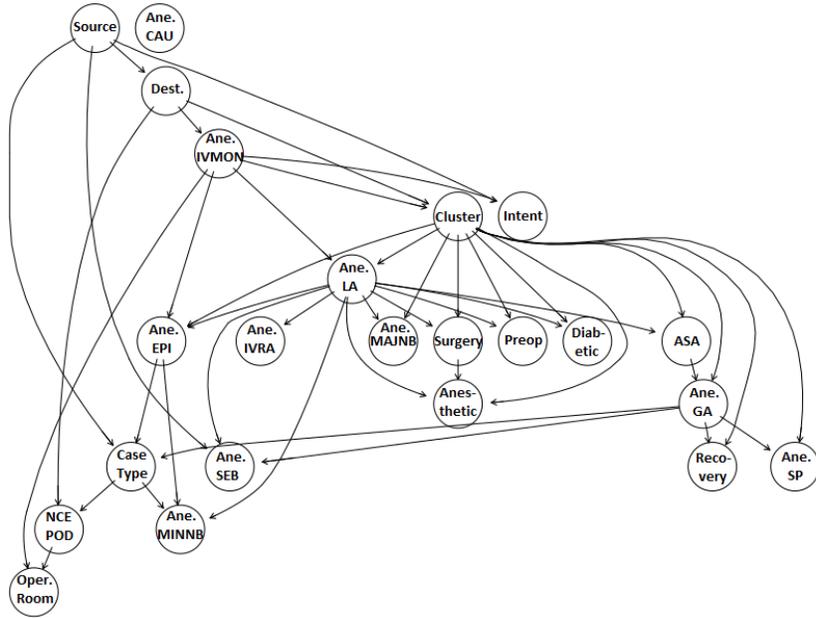


Fig. 10: A Bayesian belief network taking into account 22 attributes of a patient’s treatment. Attributes prefixed by *Ane.* denote different types of anaesthetic. Note the central role of the latent *Cluster* attribute discovered in Section 4.2. Originally published in [6].

follow the same pattern. As a baseline for comparison, results are shown for the best-fit marginal distributions reported in Table 2.

*Avoiding Data Snooping* One pitfall that was important to avoid, in particular when modelling *partial executions* of patient flows, was that of inadvertent data snooping by including the *Cluster* attribute. When evaluating a model’s predictive power on test data, the cluster should be consider an *unobserved variable*.

While intuitively obvious, this is a crucial methodological point, as including it would constitute a form of data snooping since the cluster itself is in fact derived from the target variables (cycle times). Nonetheless, the variable is able to play a role in the Bayesian network, despite being unobserved, via conditional dependencies between it, observed variables and unobserved target variables.

## 6 Discussion

The case study presented in this paper has highlighted a number of challenges and lessons learnt that can be applied to other surgical workflow modeling projects, as well wider data-driven healthcare improvement initiatives. First, data quality assurance is key. One of the most immediate observations of our analysis was presence of a good

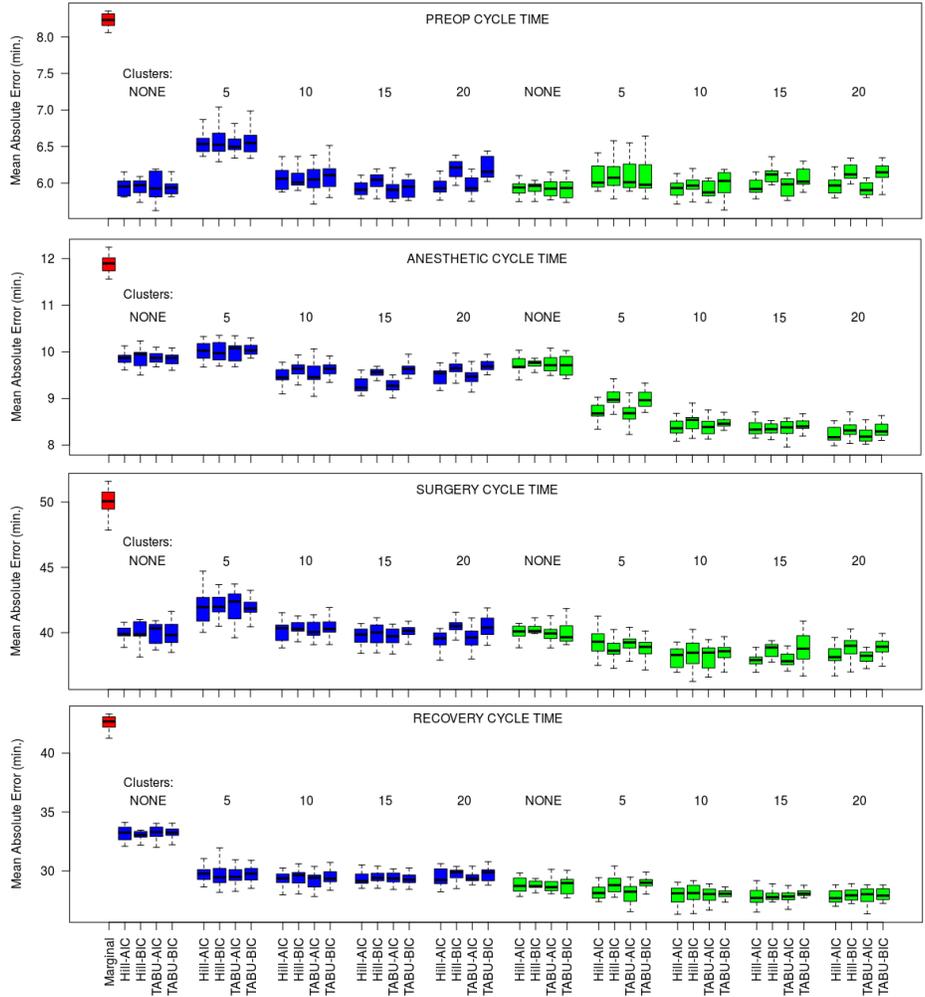


Fig. 11: Comparison of 40 different Bayesian net models using 10 (blue) and 22 (green) variables, but different structure learning algorithms. Results are based on 5 runs of 10-fold cross-validation for predicting cycle time of partially completed patient flows. As a baseline comparison, the simple best-fit marginal distribution (reported in Table 2) is shown in red. Originally published in [6].

deal of anomalous data. Process mining techniques proved to be useful for detecting outliers and identifying anomalies related to the control-flow.

Second, even though it seemed initially that there were no groupings of patients with regards to duration, PCA revealed latent patient categories in our data. Identifying these patient clusters improved Bayesian net prediction, even though interpreting what these categories mean is not straightforward.

Third, we showed that a reasonably accurate predictive model of event cycle times in the form of a simple Bayesian belief network can be built, which significantly outperforms simple marginal distribution fitting. The choice of Bayesian networks was motivated by their flexibility and interpretability, which is of great importance in safety-critical domains like healthcare. The ability to query these models suggests they would be a strong component of an intelligent probabilistic scheduling system in surgery.

Finally, Stochastic Time Petri Nets were found to be an appropriate formalism for capturing hospital policies and guidelines surrounding surgery, in particular regarding timing and resource requirements. Distinguishing between the workflows for emergency and scheduled cases was possible in a clear and transparent way, and incorporating case prioritization was straightforward. Even though the model presented in Section 5.1 is a simplified version of reality, it serves as proof-of-concept of how real-world data can be incorporated into a model that combines official procedures and guidelines with domain expert knowledge. Simulating different scenarios allowed us to test the limits of the system and to analyse the effect of varying resource allocation.

## 7 Conclusion and Future Work

In this paper, we presented a preliminary investigation into probabilistic workflow modeling, simulation and prediction methods in surgical wards. This is an important first step towards much-needed surgical care improvement. Our analysis is focused on key surgical phases, which is a level of granularity that has received less attention in existing literature.

Data-informed surgical care scheduling that takes into account individual patient characteristics, resource availability and hospital policies presents a promising approach to improving resource utilization, quality of care and, ultimately, patient and staff satisfaction. We have demonstrated the value of combining several data analysis paradigms, from mathematical modeling to process mining and machine learning, towards developing a model that effectively captures the complexity of surgical processes, while allowing for experimentation and insightful interrogation. This approach is applicable in other areas of the healthcare system, where under-utilization of expensive resources calls for precise scheduling to minimize costs and waiting times.

Our analysis considered both system-wide surgical performance (through Stochastic Time Petri Net modeling and simulation) and individual patient flow (through Bayesian Net cycle time prediction). In order to integrate the two in the future and incorporate Bayesian nets into Petri Net modeling, we propose the use of Bayesian Stochastic Petri Nets [39].

In the big data and precision medicine era, developing intelligent methods for dynamic and personalized scheduling in the surgical ward is a key research direction. Ex-

tending the work presented in this paper to incorporate more detailed information about the surgical ward is desirable, and would possibly require further domain knowledge and dimensionality reduction, so as to deal with the huge cardinality of some attributes. We also regard evaluation with domain experts as an important next step, ensuring that the recommendations of a future surgical scheduling system are understood and deemed useful by hospital staff.

## References

1. Reorganizing patient care and workflow in the operating room: a cost-effectiveness study. *Surgery* **139**(6), 717 – 728 (2006)
2. Van der Aalst, W., Weijters, T., Maruster, L.: Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering* **16**(9), 1128–1142 (2004)
3. Acid, S., de Campos, L.M., Fernández-Luna, J.M., Rodríguez, S., Rodríguez, J.M., Salcedo, J.L.: A comparison of learning algorithms for bayesian networks: a case study based on data from an emergency medical service. *Artificial Intelligence in Medicine* **30**(3), 215–232 (2004)
4. Ahmadi, S.A., Padoy, N., Rybachuk, K., Feussner, H., Heinin, S., Navab, N.: Motif discovery in OR sensor data with application to surgical workflow analysis and activity detection. In: M2CAI workshop, MICCAI, London (2009)
5. Akkerman, R., Knip, M.: Reallocation of beds to reduce waiting time for cardiac surgery. *Health care management science* **7**(2), 119–126 (2004)
6. Back, C.O., Manataki, A., Harrison, E.: Mining patient flow patterns in a surgical ward. In: HEALTHINF. pp. 273–283 (2020)
7. Baier, C., Katoen, J.P.: Principles of model checking. MIT press (2008)
8. Basin, D., Klaedtke, F., Müller, S., Zălinescu, E.: Monitoring metric first-order temporal properties. *J. ACM* **62**(2), 15:1–15:45 (May 2015)
9. Béjar Haro, B., Zappella, L., Vidal, R.: Surgical gesture classification from video data. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. pp. 34–41. Springer, Heidelberg (2012)
10. Berti, A., van Zelst, S.J., van der Aalst, W.: Process Mining for Python (PM4Py): Bridging the Gap Between Process-and Data Science. In: ICPM Demo Track (CEUR 2374) (2019)
11. Blum, T., Padoy, N., Feußner, H., Navab, N.: Workflow mining for visualization and analysis of surgeries. *Int. journal of computer assisted radiology and surgery* **3**(5), 379–386 (2008)
12. Bouarfa, L., Dankelman, J.: Workflow mining and outlier detection from clinical activity logs. *Journal of biomedical informatics* **45**(6), 1185–1190 (2012)
13. Bouarfa, L., Jonker, P.P., Dankelman, J.: Discovery of high-level tasks in the operating room. *Journal of biomedical informatics* **44**(3), 455–462 (2011)
14. Bucci, G., Carnevali, L., Ridi, L., Vicario, E.: Oris: a tool for modeling, verification and evaluation of real-time systems. *International journal on software tools for technology transfer* **12**(5), 391–403 (2010)
15. Burattin, A., Maggi, F., Sperduti, A.: Conformance checking based on multi-perspective declarative process models. *Expert Systems with Applications* **65** (03 2015). <https://doi.org/10.1016/j.eswa.2016.08.040>
16. Cochran, J.K., Bharti, A.: Stochastic bed balancing of an obstetrics hospital. *Health care management science* **9**(1), 31–45 (2006)
17. Denton, B., Viapiano, J., Vogl, A.: Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health care management science* **10**(1), 13–24 (2007)

18. Fages, F., Rizk, A.: From model-checking to temporal logic constraint solving. In: International Conference on Principles and Practice of Constraint Programming. pp. 319–334. Springer (2009)
19. Fu, J., Topcu, U.: Computational methods for stochastic control with metric interval temporal logic specifications. In: 2015 54th IEEE Conference on Decision and Control (CDC). pp. 7440–7447. IEEE (2015)
20. Funkner, A.A., Yakovlev, A.N., Kovalchuk, S.V.: Towards evolutionary discovery of typical clinical pathways in electronic health records. *Procedia computer science* **119**, 234–244 (2017)
21. Huang, Z., Lu, X., Duan, H., Fan, W.: Summarizing clinical pathways from event logs. *Journal of Biomedical Informatics* **46**(1), 111–127 (2013)
22. Hulshof, P.J., Boucherie, R.J., Hans, E.W., Hurink, J.L.: Tactical resource allocation and elective patient admission planning in care processes. *Health care management science* **16**(2), 152–166 (2013)
23. Jiménez Ramírez, A., Barba, I., Fernandez-Olivares, J., Del Valle, C., Weber, B.: Time prediction on multi-perspective declarative business processes. *Knowledge and Information Systems* **57**(3), 655–684 (2018). <https://doi.org/10.1007/s10115-018-1180-3>
24. Kayis, E., Wang, H., Patel, M., Gonzalez, T., Jain, S., Ramamurthi, R., Santos, C., Singhal, S., Suermondt, J., Sylvester, K.: Improving prediction of surgery duration using operational and temporal factors. In: AMIA Annual Symposium Proceedings. vol. 2012, p. 456. American Medical Informatics Association (2012)
25. Koller, D., Friedman, N.: Probabilistic graphical models: principles and techniques. MIT press (2009)
26. Lalys, F., Jannin, P.: Surgical process modelling: a review. *International Journal of Computer Assisted Radiology and Surgery* **9**(3), 495–511 (2014)
27. Lin, H.C., Shafran, I., Yuh, D., Hager, G.D.: Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions. *Computer Aided Surgery* **11**(5), 220–230 (2006)
28. Mans, R., Reijers, H., van Genuchten, M., Wismeijer, D.: Mining processes in dentistry. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium. pp. 379–388. ACM (2012)
29. Martina, S., Paolieri, M., Papini, T., Vicario, E.: Performance evaluation of fischer’s protocol through steady-state analysis of markov regenerative processes. In: 2016 IEEE 24th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS). pp. 355–360. IEEE (2016)
30. NCEPOD: NCEPOD classification of intervention. <https://www.ncepod.org.uk/classification.html> (2019), accessed: 2019-11-22
31. Neumuth, T., Jannin, P., Schlomberg, J., Meixensberger, J., Wiedemann, P., Burgert, O.: Analysis of surgical intervention populations using generic surgical process models. *Int. Journal of Computer Assisted Radiology and Surgery* **6**(1), 59–71 (2011)
32. NHS Scotland: National theatres project report. <https://www.isdscotland.org/Health-Topics/Quality-Indicators/National-Benchmarking-Project/National-Theatres-Project/> (2006), accessed: 2019-11-22
33. Paolieri, M., Horvath, A., Vicario, E.: Probabilistic model checking of regenerative concurrent systems. *IEEE Transactions on Software Engineering* **42**(2), 153–169 (2015)
34. Royal College of Anaesthetists: Perioperative medicine: the pathway to better surgical care (2015), London
35. Scutari, M.: Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software* **35**(3), 1–22 (2010), <http://www.jstatsoft.org/v35/i03/>

36. Stahl, J.E., Rattner, D., Wiklund, R., Lester, J., Beinfeld, M., Gazelle, G.S.: Reorganizing patient care and workflow in the operating room: a cost-effectiveness study. *Surgery* **139**(6), 717–728 (2006)
37. Stauder, R., Okur, A., Peter, L., Schneider, A., Kranzfelder, M., Feussner, H., Navab, N.: Random forests for phase detection in surgical workflow analysis. In: *Int. Conf. on Info. Processing in Computer-Assisted Interventions*. pp. 148–157. Springer (2014)
38. Strum, D.P., May, J.H., Vargas, L.G.: Modeling the uncertainty of surgical procedure times: comparison of log-normal and normal models. *Anesthesiology* **92**(4), 1160–1167 (2000)
39. Taleb-Berrouane, M., Khan, F., Amyotte, P.: Bayesian Stochastic Petri Nets (BSPN) - a new modelling tool for dynamic safety and reliability analysis. *Reliability Engineering & System Safety* **193**, 106587 (2020)
40. Westergaard, M., Maggi, F.M.: Looking into the future: Using timed automata to provide a priori advice about timed declarative process models. In: Meersman, R., Panetto, H., Dillon, T., Rinderle-Ma, S., Dadam, P., Zhou, X., Pearson, S., Ferscha, A., Bergamaschi, S., Cruz, I.F. (eds.) *OTM 2012*. pp. 250–267. Springer, Heidelberg (2012). [https://doi.org/10.1007/978-3-642-33606-5\\_16](https://doi.org/10.1007/978-3-642-33606-5_16)