IWISH - Intelligent Workflow optimization and
Intuitive System interaction in Healthcare

DELIVERABLE D3.1
State of the Art description of clinical procedure tracking

Project number:	AI2021-066
Document version no.:	v1.0
Edited by:	John van den Dobbelsteen
Date:	22-12-2022

Eureka AI Thematic call

HISTORY

Document version #	Date	Remarks
V0.1	1 Nov 2022	Starting version, template
V0.2	1 Dec 2022	Compilation of first input by partners
V1.0	22 Dec 2022	Final version

Deliverable review procedure:

- **2 weeks before due date**: deliverable owner sends deliverable –approved by WP leader– to Project Manager
- **Upfront** PM assigns a co-reviewer from the PMT group to cross check the deliverable
- **1 week before due date**: co-reviewer provides input to deliverable owner
- **Due date:** deliverable owner sends the final version of the deliverable to PM and co-reviewer

TABLE OF CONTENTS

1 Introduction

1.1 Aim of the activity

The goal of this document is to provide a description of the state-of-the-art in modelling approaches for tracking the progress of clinical procedures. In the complex environment of the OR, effective use is limited by precision in scheduling elective surgical cases and interventions, which can only be done optimally when there are reliable predictions of the time these cases will consume. Models of surgical progress can provide improvements in estimations of surgical durations, hereby increasing OR efficiency. Although it is known that surgical duration is determined by a broad range of factors, the current methods of OR planning are often based only on either average surgery durations or estimates by the surgical staff, leading to inconsistencies between planned and actual surgery durations. Surgery duration estimation can be performed by both offline models, using pre-operative data and online models, mainly using intra-operative data. The on-line models are preferable, due to their inherent capability to deal with unexpected events, hereby providing greater promise for improving OR efficiency and workflow.

The descriptions are based on open literature and highlight the current forefront in surgical duration models. The following aspects of the estimation models are distinguished:

- *Model:* First, we describe the type of model which amongst other things includes the granularity of the prediction. Here one can also incorporate the extent of clinical applicability of the model. This describes the breath of procedures to which the model is applicable.
- *Acquisition:* This aspect describes the data sources that are used as an input for the model. These could for example be patient characteristics (pre-operative) or live data from the OR (intra-operative).
- *Analysis:* The data sources can be analyzed using different types of pattern recognition and machine learning techniques, for example binary classifiers, regression methods and artificial neural networks.
- *Validation and Performance:* To compare relative study quality the validation of the model is also analyzed, for example whether the authors used separate test and training data and the size of these data sets.

1.2 Glossary

ANN	Artificial Neural Networks
CART	Classification and regression trees
DTW	Dynamic Time Warping
HMM	Hidden Markov Model
LLR	Log-linear regression
LR	Linear regression
MARS	Multivariate adaptive regression splines
MCMC	Monte Carlo Markov Chain
OR	Operating Room
RF	Random Forest
SVM	Support vector machines

2 Online estimation of surgical progress

2.1 Model

In terms of modelling the surgical duration, two major approaches can be identified. The most common and extensively studied method involves a prediction during *pre- operative planning*. These methods use only information available prior to the actual surgery to estimate the duration of the surgery. This information is then used allocate the right amount of time to the elective surgery in the operating room planning.
The second approach (also) uses *intra-operative information* and is therefore usable as an online model of surgery duration. These models can cope with unexpected delays and complications and update the duration estimates in a live setting. It needs to be noted that most studies that cover online models, do not provide a fully automated solution. As we will see in the following parts of this review, the type of data sources associated with live intra-operative data are often unstructured and messy, making a fully automated model a challenging task.

2.1.1 Clinical application

Considering the clinical application of the surgery duration models we can again identify several categories (Table 2.1). The broadest type of model is applicable to all the procedures available in the data set of the hospital (n=23). Other researchers developed models to predict the duration of a surgery in a specific medical specialty, or group of procedures (n=7). Finally, models are developed to assess specific medical procedures (n=24). The most notable procedure in the latter category is the laparoscopic cholecystectomy, accounting for almost half of the identified research articles. Considering the previously made distinction between online and offline models, it should be noted that all online models in previous research (n=19) are applicable only to specific procedures. None of the research included in this review has provided an online model with a broader clinical application, for example to multiple procedures within a medical specialty.

2.2 Acquisition

The acquisition aspect of the surgery duration estimation system deals with the variables that are used as an input to the model. In choosing the right variables there are several aspects to consider. First and foremost, the variable should have predictive value for the duration of the surgery. If there is no correlation between the identified variable and the duration of a surgical case, there is no point in including the variable into the model.
The following sections will highlight some research in identifying important factors, both in offline models, using pre-operative data and online models, using intra-operative data or a combination between pre-operative and intra-operative data. Other aspects to consider are the measurement level and availability of the variable. Variables on a high measurement level (i.e. continuous variables on a ratio scale) are often most informative and yield high predictive value. Naturally, variables also need to be readily available for (automatic) measurement.

Category	Details	Number of studies	References
All	All available surgeries	23	[17, 77, 41, 71, 20, 40, 22, 51, 81, 83, 54, 97, 86, 96, 95, 70, 80, 32, 19, 1, 83]
Medical specialty	Cataract- and Oculoplastic surgery	1	[16]
	Colorectal surgery	1	[90]
	Endoscopic surgery	1	[14]
	Orthopedic surgery and General surgery	2	[49, 74]
	Thoracic surgery	1	[18]
	Trauma surgery	1	[9]
	Vascular surgery	1	[43]
Specific procedure	Brain tumor removal	3	[27, 64, 28]
	Laparoscopic cholecystectomy	11	[35, 4, 2, 38, 12, 79, 89, 11, 66, 10, 52]
	Laparoscopic myomectomy	1	[36]
	Laparoscopic ovarian endometrioma	1	[30]
	Lumbar discectomy	3	[27, 55, 26]
	Robot-assisted hysterectomy	1	[56]
	Robot-assisted radical prostatectomy	1	[3]
	Robotic endoscopic coronary artery bypass	1	[92]
	Third molar removal	1	[87]

Table 2.1: The clinical application of the surgery duration models can be separated into models applicable to all surgeries (n=23), models applicable to one or more medical specialties (n=7) or models applicable only to a specific medical procedure (n=24).

2.2.1 Pre-operative data

Online forecasting of surgery durations needs intra-operative data, which is data that is recorded and analyzed in (near) real-time during the actual surgery. The distinct data sources can be characterized as apparatus use, medical instrument usage and location, patient monitoring, surgeon activity (including hand monitoring) and video (Table 2.2).

Category	Variables	Number of studies	References
Apparatus use	Usage of neuro-navigation device, ultrasound device, and neurophysiology monitor	1	[28]
Instrument use and location	(Binary) usage patterns of medical instruments, monitoring using RFID tags	11	[1, 2, 11, 12, 27, 28, 56, 55, 64, 66, 79]
Patient monitoring	Pulse oximetry and vital signs monitors	1	[1]
Surgeon activity (including hand monitoring)	Hand kinematics, gaze-tracking	5	[1, 27, 26, 38, 52]
Video	Laparoscopic or microscopic video recordings or recordings from the complete OR	4	[9, 10, 88, 89]

Table 2.2: Categorized overview of variables that have an important influence on surgery duration using *intra-operative* modelling.

The following sections will give a summary of data used in intra-operative data to predict surgery duration. Some of the reported studies did not specifically compute surgical case durations, but instead aimed to construct a model of the surgery, a 'surgical process model'. Since these models can among other things be used to predict the duration of a surgery, said studies have been included in this review.
An important aspect to note in assessing studies using intra-operative data is how the data is acquired. Several studies use manual acquisition or labelling of the data. For example, Ahmadi et al. (2006), used manual classification of the state of the surgery from video recordings, meaning that someone had to watch the whole dataset to apply the correct labelling. Automatic video processing, however crude, is therefore a striking advantage in these kinds of studies, if the aim is to use the prediction system in practice at some point. This point will be further elaborated on in the next chapter.

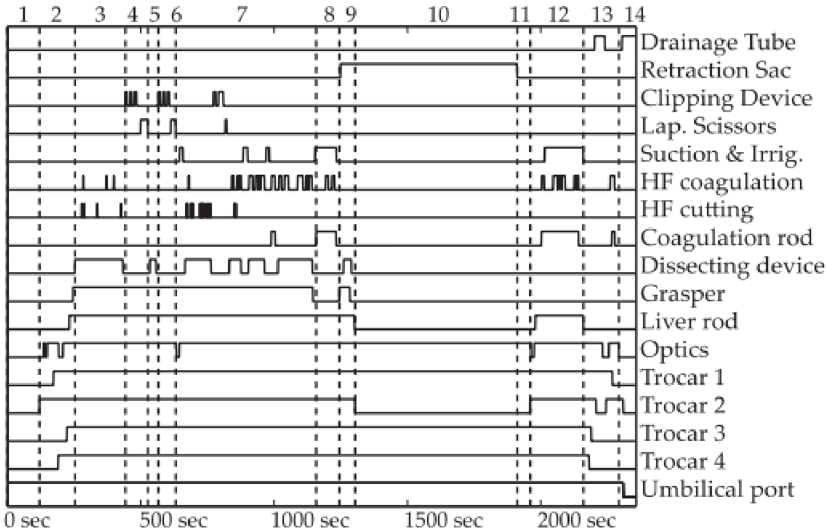

Figure 2.1: Binary usage data of surgical tools is the most popular data sources for online models of surgical duration and surgical phase prediction. The figure from a study by Padoy et al. (2012) shows an example of 18 binary time series indicating the use of medical instruments during a laparoscopic cholecystectomy [66]. It needs to be noted that in this model, and many other similar developed models, the medical instrument usage was labelled manually.

Apparatus use

Next to medical instruments, which will be discussed in the next section, it is also possible to use the usage of devices present in the OR. In a study on sixty neurosurgical cases, Franke et al. (2015) created a model using, amongst other sources, information on usage of apparatus present in the operating room [28]. The neuro-navigation device, ultrasound device, and neurophysiology monitor were classified as ”not used yet”, ”in use”, ”likely to be required”, ”not likely to be required”, ”not required”, and ”unused”. Based on the assigned state, this information was used to predict surgical phases.

Instrument use

As a surgery progresses through different phases, the surgeon often uses a specific set of tools. Hence, tracking the usage of medical instruments has become a popular approach when creating online models for surgery duration estimation.

Ahmadi et al. (2006) [2] predicted the states of a laparoscopic cholecystectomy using usage information of all 14 instruments. A model using 17-binary inputs indicating the use of each instrument was able to detect changes in 14 surgical phases within a range of 5 seconds in 92% of the cases. An example of these binary tool usage signals is given in Figure 3.1. Instruments were weighed according to their synchronisation across different surgeries, giving higher weight to instruments that were used consistently over many surgeries. Unfortunately, the authors do not report which instruments are most relevant. It needs to be noted that the instrument use was extracted from OR video using manual labelling, so in the current form the model is not suitable for real-time use.

Using RFID technology, Agarwal et al. (2007), created a system that could track the location of medical instruments in the operating theatre [1]. Fusing this information with tracking of staff location and drug location, a low-level event record was created. Based on a set of seventeen pre-defined rules, a medical encounter record was generated, listing medically relevant events, which could for example be the surgery nearing the end.

Padoy et al. (2008) [65] used real-time monitoring of endoscopic camera and instrument use to detect in which of the 14 phases of laparoscopic cholecystectomy the case is proceeding. The model was cross-validated on 11 laparoscopic cholecystectomies and provided a detection rate of 93% and according to the authors could reliable identify relevant events (such as the end of surgery time). Again, the instrument use was labelled manually on the recordings. In a follow-up study by the same research group, Padoy et al. (2012), used Dynamic Time Warping (DTW) and Hidden Markov Models (HMM)to predict phases in laparoscopic cholecystectomies [66]. The online models reached an accuracy of over 90% and the model could predict remaining surgery time with mean prediction error below 5 min, when the surgery was in the tenth of fourteen phases. Other research by the same research group used similar approaches in predicting surgical phases of laparascopic cholecystectomy using binary tool usage data, with minor adjustments to model and data recording setup (e.g. [11]).

Bouarfa et al. (2011) [12] used a Bayesian Hidden Markov Model to detect high-level surgical tasks based on low-level sensor data in laparoscopic cholecystectomy. The algorithm could detect high-level tasks (i.e. phases in the surgical procedure, for example ”clipping and dissection” or ”gallbladder removal”) with 90% accuracy when using noise-free sensory inputs. The sensory data used included binary variables indicating instrument use, which were manually generated from video streams. Simulation of signal noise resulted in significant decrease in model accuracy.
Nakamura et al. (2013) predicted duration of brain tumor resection using mean removal speed [64]. By tracking the instrument location in real-time and combining this information with previously recorded MRI-scans of the tumor size and location, the algorithm was able to compute surgery process, with an average error of 14 minutes (± 9 minutes standard deviation) over the whole surgery, declining towards the end.

In a study on two surgical procedures (lumbar discectomy and brain tumor resections), Franke et al. (2013), proposed a model based on low-level surgical tasks to predict intervention time [27]. The surgical task was defined as a combination of an actor, activity, instrument, anatomical structure and intervention phase. For example, the nurse disinfecting the skin using a swab during preparation. The model was able to predict surgery duration with an error of around 10 minutes for the discectomy and around 15 minutes for brain tumor removal. However, the quintuple task variables were annotated manually by human observers, rendering the current model unsuitable for automated use. A recent study by the same research group included patient status and device usage to further improve the model on predicting phases in brain tumor removal [28]. However, still human observers were needed to generate input data.

Stauder et al. (2014) [79] detected surgical phases of laparoscopic cholecystectomy using a random forest (RF) model based on instrument usage data and various other data sources. An accuracy of 65% was obtained when using seven distinct surgical phases, this improved to over 85% for three surgical phases. The authors noted that even features that appear to carry very little information can have a high impact on the classification, such as CO_2 pressure in the abdominal cavity, suction bag weight and whether the surgical light is switched on. In a follow-up study by the same group (Stauder et al., 2014b [78]), again a Random Forest (RF) model was used on laparoscopic cholecystectomy to detect one of seven surgical phases. Based on intraabdominal pressure, suction and irrigation bag weights, table inclination and binary data of tool use, acquired using RFID tags, phases were detected with a precision of 87.6% and a sensitivity of 75.4%.

In a study on forty lumbar discectomies, Maktabi et al. (2015) estimated surgery duration using frequency domain analysis of surgical activity time-series [55]. A total of 35 operational (instruments used), spatial (treated body parts) and organizational (executing person) binary time series was generated. After transformation to frequency domain, signal features were used to assess surgical duration, which the best signals achieving around ¿20% error. The time series were manually generated making the current method unsuitable for online use.
Malpani et al. (2016) [56] detected phases in a robot-assisted hysterectomy using system events. Using information from the Da Vinci surgical robot, the use of tools and the built-in camera could be recorded automatically. In a set of 24 surgeries, the model was able to detect surgical phases with an accuracy in the range from 66%-76% using three different classifiers.

In a recent study, Guedon et al. (2016) [34] estimated elective laparoscopic cholecystectomy duration in real-time using information from an electrosurgical device. Several features of the electrosurgical device activation pattern were extracted, including the first and last time of activation, the number of activations and the total duration of the acti- vation, which all positively affected classifier performance. Various pre-operative features (including patient age, BMI and surgeon identifier) were tested but did not increase model accuracy.

Patient monitoring
In their 2007 study, Agarwal et al. suggest using patient monitoring to detect the status of the patient during the surgery [1]. They report using data streams from pulse oximetry and vital signs monitors tracking heart rate and blood pressure. It is not reported how the data is retrieved from the patient monitoring systems to be used in the model, or how important the patient data turned out to be for coming to proper predictions.

Surgeon activity and hand tracking
Instead of tracking the patient or solely the instruments, an approach can be to track the activity of the surgical team. In their previously described tracking system, Agarwal et al. (2007), equipped the surgeon and nursing staff with RFID tags to track their location [1].
Based on eye movements of the surgeon, James et al. (2007) [38], developed a model to recognise phases in a porcine laparoscopic cholecystectomy. The eye-gaze data contains information of underlying surgical activity and an accuracy of 66% was reported using an artificial neural network (ANN) model. This improved to 75% by adding data relating to the instrument use, like the binary usage signals described before.
Loukas & Georgiou (2013) used hand kinematics as an input for a model predicting surgical phases [52]. The hand movements of the surgeon were tracked by placing orientation sensors on the instruments of a Virtual Reality simulator for laparoscopic training. A precision between 59% and 91% was achieved for the distinct phases of a VR simulated cholecystectomy surgery.
Forestier et al. (2015) applied a decision tree model on a data set of 22 lumbar disc herniation surgeries. The input data used considered of one data triplet per hand of the surgeon, which consisted of the action, anatomical structure and instrument. For

example: (cut, muscle, scissors). The triplets were manually labelled by a human observer, sensory recordings were simulated by adding noise to the manually generated labels.

Video

A final category of OR signals is video, which stands apart a bit from the other categories as it is a lower level of measurement. In other words, OR video or endoscopic video is recorded to infer information on patient status or tool usage. However, since the specific types of challenges and approaches related to video, it is discussed here as a separate category.

Bhatia et al. (2007) estimated the OR state (one of ’empty’, ’transitioning’ or ’in-use’) using video. The states could be accurately predicted in real-time (1 second) from OR video. Note that the authors did not give a prediction, but instead a real-time indication of what was happening in the OR. [9] Lalys et al. (2012) [47] predicted phases of cataract surgery using only microscopic video data. The miscrocopic video was automatically analyzed using shape, colour, texture and other features and processed using Hidden Markov Models and Dynamic Time Warping. These visual cues could provide an phase identification accuracy of 91% FRR for HMM and 94% FRR for DTW. The FRR is the frequency recognition rate, or the fraction of video frames that was classified correctly. Twinanda et al. (2016) [89] studied task recognition on laparoscopic cholecystectomy procedures using ”EndoNet”, a convolutional neural network (CNN). Based on laparascopic video, the CNN automatically extracts relevant features. These are first used as an input to an support vector machine (SVM) model, which output is in turn used to detect surgical phases using a Hidden Markov Model. The reported overall accuracy was 92% using an offline model and 81% using an online model. A novel approach using video data to estimate surgical phases was used by Tran et al. (2016) [88], by retrieving optical flow vectors from video. Optical flow describes the pattern of motion of objects, surfaces, and edges between two video frames. The retrieved vectors are then simplified into four directions (up, down, left, right). The optical flow vectors are used as an input to Latent Dirichlet Allocation (LDA) and Hidden Markov Models (HMM), where the best model yielded 73% accuracy.

2.3 Analysis

An extensive range of machine learning and pattern recognition techniques can be applied to the field of surgical duration estimation (Table 2.3). The following section provides a concise introduction to these methods and their respective advantages and limitations.

Category	Number of studies	References
ANFIS	1	[16]
Artifical neural network (ANN)	3	[89, 16, 56]
ANOVA	1	[80]
Bayesian methods	2	[17, 19]
CCA, PCA	1	[10]
Decision trees (CART) and	5	[26, 32, 71, 56, 79]

Random forest (RF)		
Dynamic Time Warping (DTW)	3	[2, 10, 66]
Fuzzy Logic (Expert system)	1	[1]
GMMAR	1	[52]
Hidden Markov Model (HMM) and Markov Chain Monte Carlo (MCMC)	6	[10, 11, 12, 28, 66, 88]
K-means clustering	1	[88]
Latent Dirichlet allocation	1	[88]
Linear regression and log-linear regression	17	[64, 87, 40, 43, 90, 22, 51, 35, 74, 83, 95, 70, 36, 16, 51, 32, 71]
Logistic regression	3	[4, 3, 30]
M5	1	[32]
Multivariate adaptive regression splines (MARS)	1	[71]
Mean and median	4	[49, 18, 54, 97]
Multiplicative regression	1	[41]
Pearson correlation	1	[92]
Power Spectral density (Frequency domain analysis)	1	[55]
Probability distribution	8	[19, 20, 81, 85, 83, 86, 77, 96]
Support Vector Machines (SVM)	3	[9, 89, 56]

Table 2.3: A wide variety of techniques from the field of pattern recognition and machine learning have been used in surgery estimation models. The most popular being (log-)linear regression (n=17) and simple probability distributions (n=8), which are both mostly used for offline models. For online models, the Markov models, such has the Hidden Markov Model (HMM) are most popular (n=6), together with Dynamic Time Warping (DTW, n=3), a technique originating in speech recognition.

2.3.1 Regression and splines

The advantages of linear regression are that the model is simple and linear in the parameters, so it can be fitted algebraically using the linear least-squares method, rendering the method computationally fast. The influence of the independent variables on the dependent variable can be observed from the regression equation, so the model can be easily interpreted. A downside of the linear regression model is that it assumes a linear relation between the input variables and output, which is of course not always the case. A variant, the log-linear model, is obtained by predicting the log-transformation of the dependent variable. It can be argued that this is a suitable model for surgery duration estimation by noting that the underlying data is described accurately by a log-normal distribution.

Another extension of the model yields the mixed effects linear model, which uses both fixed and random effects. This model was for example used by Eijkemans et al. (2010), who used the type of operation as a random effect and the other variables as fixed effects [22]. In practice this means the coefficient of the random effect is conditional, in this case on the operation type, whereas for the fixed effects the coefficient multiplied with the independent variable is constant for the whole set.

Some dependent variables exhibit non-linear relationships with the independent variables, which causes poor fit of a regular linear regression model. Splines are a possible to solution to this problem. Splines are continuous functions formed by connecting a series of similar basis functions. The points where the segments are connected are called knots or hinges.
A popular variant of splines are the Multivariate adaptive regression splines (MARS), a method first coined by Friedman (1991) [29]. The MARS model is defined using linear splines, but splines can of course use higher order basis functions. ShahabiKargar et al. (2014) used the MARS model to predict surgery duration, highlighting the advantages that MARS can search a large number of variables and their possible non-linear interactions [71].

2.3.2 Decision Trees and Random Forests

Decision trees can be subdivided in classification and regression trees (CART), the main difference being the classification trees provide a categorical output, where regression trees provide a continuous estimate of the dependent variable. A decision tree can be visualized as a graph, where each node poses a certain question (e.g., x1 < 5). Depending on the answer, the correct edge is chosen leading to another question at the following node. This finally leads to either a categorical or numerical prediction of the outcome variable. As these models are based on a defined set of rules, they can easily be understood and interpreted. The structure of the model allows for non-linear relationships.

An extension of decision trees are Random Forests, which have been used for the prediction of surgery durations [71, 79]. As the name suggests, the forest consists of a collection of trees, which are each trained on a random subset of the training set. The outcome of each decision tree counts as a vote for a certain outcome. The modal (in case of classification) or mean (in case of regression) outcome of all trees provides the final prediction. Random Forests usually perform better than decision trees, but this comes at the expense of interpretability. Because of the number of trees (and hence different models) used, random forests are able to deal well with situations that have missing data. Another advantage is that the model can be used to assess the importance of certain variables, by excluding them from certain permutations and observing the loss of accuracy [79].

2.3.3 Support Vector Machines (SVM)

Support Vector Machines (SVM) is a widely used technique for classification, which has also been applied to surgery duration estimation (e.g. [34, 9]. In principle, SVM is a binary classifier that is trained to have a linear decision surface. The decision surface is constructed in such a way that it provides the largest distance between two sets of data points, belonging to different classes. By using so called kernels, SVM can be extended to accommodate non- linear relations. A limitation SVM is that it is principly used for binary classification, rather than regression.

2.3.4 Markov Models (MCMC, HMM)

A discrete-time Markov-chain is a mathematical representation for a series of events. The representation consists of a certain number of states, that are linked by probabilities of transferring from one state to another. An important characteristic of a Markov process is that the probability of moving from one state to the next is only defined by the current state of the system, not by any other state in the history of the event-chain, the system is so called ”memory-less”. Markov models are specifically useful in modelling time series that consist of specified events.

In a Markov Chain Monte Carlo (MCMC model, the state-transition probabilities are estimated using many random simulations. Luangkesorn & Eren-Dogu (2016) used MCMC on surgery duration estimation and found good results, even in cases with a low sample sizes [53].
Hidden Markov Models (HMM) is an extension of the Markov chain model which makes a distinction between observable states and hidden states. In terms of surgery phase recog- nition, the real phases (e.g. suturing) might be unobservable, but observable by the visible outputs (e.g. use of a specific instrument). Bhatia et al. (2007) used a HMM to predict which of four states a surgery was in, (Figure 2.2) [9]. The likeliest path based on the Markov-chain structure and observed outputs can be found using the Viterbi-algorithm [47]. A Markov model requires an explicit description of the system that needs to be modelled, i.e. in surgery duration estimation, the distinct phases of the surgery need to be defined. This can limit the applicability of the model to other clinical applications. Lalys et al. (2012) suggest specific pre-processing steps to make the inputs more uniform and as such make the system more adaptable.

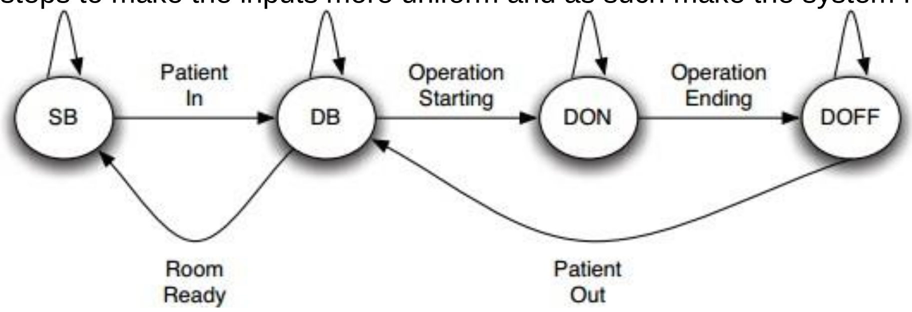

Figure 2.2: An example of a Markov Chain model used for the identification of phases in a surgery. The four phases refer to four possible states the surgery can be in (SB,DB,DON,DOFF). During the training of the Markov model, the arrows between the stages all get weighted based on the input. For each input, there is then a probability to change states and a probability to remain in the same state. Image retrieved from Bhatia et al. (2007) [9]

2.3.5 Dynamic Time Warping (DTW)

Dynamic Time Warping is an algorithm used to detect similarities in time-series data [62]. Using two time series A(t) and B(t), a txt distance matrix is constructed that contains the distance between A and B at any given point in time. Finding the minimal values in the distance matrix then allows for alignment of the time series (dynamic warping) in such a way that they are most similar. In the case of surgical process modelling, DTW can be used to compare the current procedure to several classes of procedures and classifying the procedure to the one that is most similar.

Compared to Markov-chain models the upside of DTW is that it does not need to have an explicit model of the surgery [2]. Limitations of DTW include the fact that it can only handle linear sequences of events, where Markov chains allow for bifurcations [11]. DTW also needs the complete procedure to be completed before the classification can be done, rendering the method unsuitable for real-time classification purposes [34].

2.3.6 Artificial Neural Networks (ANN)

Artificial neural networks comprise a whole set of models, that are inspired by the functionality of the brain. The model typically consists of input layers, hidden layers and output layers. During the training phase, connections between these neurons are made, which are given certain weights, either positive (excitatory) or negative (inhibitory). Research into ANN has been enormous of the past decades and the

specific implementations and varia- tions of ANN go beyond the scope of this report. In general, it can be stated that ANN are able to capture complex, non-linear relationships and generally provide good results. However, ANN are black-box models and are very hard to interpret.
In literature on surgical duration estimation, several authors have used artificial neural networks. James et al. (2007) used a Parallel Layer Perceptron (PLP) topology to recognize progress in a porcine laparoscopic cholecystectomy using eye-gaze data [38]. Devi et al. (2012) used standard artificial neural networks and adaptive neuro-fuzzy inference systems (ANFIS) to predict surgery durations in an ophthalmology department [16].

2.4 Validation and Performance

In the previous sections we have seen many different models, applications, data sources and processing methods. Here a few notes on performance aspects are made.

2.4.1 Data set split and size

One of the most important best practices in modelling is out-of-sample validation, also known as separation of test and training sets. First, the model should be trained on the so-called training set. Validation of the model should be performed on a set of unseen test data, to minimize the risk of overfitting the model to the training data, as highlighted in figure 3.4 Several of the cited studies are unclear on whether separate test and training data sets have been used for model validation (e.g. [71, 36]).

Generally, increasing the size of the data set will generate better models. In terms of surgery duration estimation, the obvious unit to consider is the number of cases. However, when comparing the data set size in terms of number of cases used to generate the model, there is one important consideration to make, which is the amount of data that is recorded per case. Consider for example an offline model, that uses some patient characteristics (e.g. age, BMI and comorbidities) to predict case duration. A data set of thousands of these cases with pre-operative date might constitute less actual information than the data used to train an online prediction model using the OR video data of only a few surgical cases.

2.4.2 Performance

Secondly the choice of a performance metric is important. Several metrics can be identified to describe the error. In case of a numerical prediction, a commonly reported metric is the root mean squared error (RMSE). In the case of surgery durations this commonly refers to the amount of minutes that the prediction is off compared to the real duration of the surgery. Such a performance metric makes it hard to compare the errors on surgical cases with different lengths, as it can be imagined that a 5 minute error on a short surgery is of way more importance that the same absolute error on a surgical cases taking several hours. It is therefore advisable to include relative metrics such as the mean absolute percentage error (MAPE).

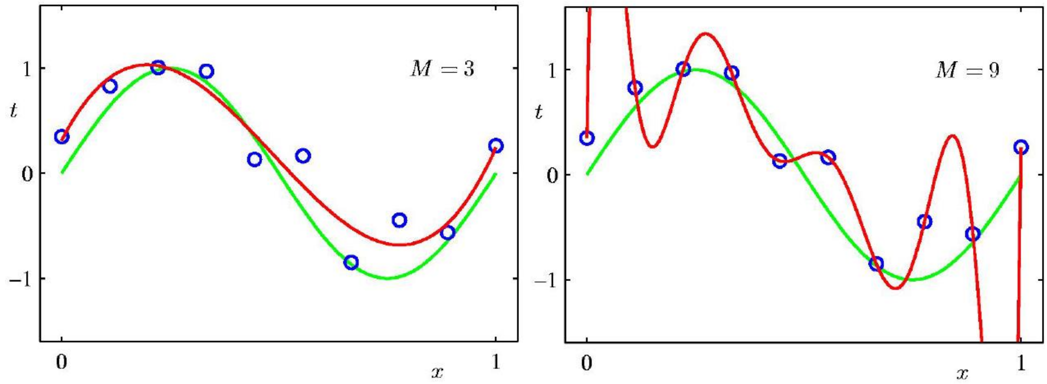

Figure 2.3: A serious risk in modelling is to overfit the data to the training sample, making the model perform badly on unseen data (a test set). On the left a third order fitted polynomial is shown (red line). The model fits the data with a certain error, but it will perform about equally well if you would generate ten points from different parts of the sine wave. In other words, the training and test set errors would be comparable. On the right side, a ninth order polynomial function is fitted to the data. The training set is fitted perfectly, but the model will have a far greater error on unseen test data, making this a definite case of overfitting. Faisal (2015) [25]

When the model has a categorical output, for example by classifying one of several states of a surgery, other performance metrics are used. In case of a binary classifier, e.g. predicting yes or no for the surgery being in a certain states, four outcomes are possible: true positives, true negatives, false positives and false negatives. Specificity (SPC) is also known as false positive rate. The sensitivity (SEN) is sometimes called true positive rate or recall. Finally, the accuracy (ACC) gives the fraction between false and true predictions of the model.

3 Conclusions

Based on the information retrieved during the literature review, some general insights on the state-of-the-art in surgery duration estimation can be formulated.

3.1 Characteristics of surgical cases

Several authors have researched the underlying distribution of surgical procedure times. Identifying the correct distribution can be used as a crude predictive model, as the probability distribution describes the chance that a certain surgery time occurs. The most used distributions are the normal and log-normal distribution. On a data set of over 40,000 cases Strum et al. (2000) concluded that the lognormal distribution provides are better fit compared to the normal distribution, a finding that was similar to research on the prediction bounds of the lognormal model by Zhou and Dexter (1998) [83, 96]. May et al. (2000) compared the normal distribution to the two-parameter and three-parameter lognormal distributions on a sample of 60,000 surgical cases [57]. They hypothesized that the addition of an additional shift or location parameter would improve the goodness of fit of the lognormal model, as surgical procedures always have durations larger than zero. It was verified that the lognormal model provided better fit compared to the normal model and especially in cases with high positive skew the three-parameter lognormal model should be used.

3.2 Characteristics of surgical duration models

From the reviewed literature, several important properties of a model can be identified. First, a broadly usable model needs to be able to handle procedures with few samples of historic data [18]. An often-used work-around for handling procedures with few historic samples is to assume a certain probability distribution of case durations (see previous section). As the log-normal distribution can be described by two parameters, a prediction bound could in theory be generated having only two historical cases [96]. In the case where the specific surgeon has not recently completed a similar surgery, Macario and Dexter (1999) suggest simply taking the mean of historical cases by other surgeons performing the same procedure, as this gives similar results as more sophisticated statistical methods [54]

Bayesian methods are a promising approach to deal with data that has little historic cases as it provides a means to combine historic data (prior) with expert opinion to form an improved estimate of the case duration (posterior). Stepaniak et al. (2009) concluded that using a Bayesian method incorporating the prior estimates of surgeons improved duration estimation when less then ten historic case durations were available [81]. Dexter et al. (2013) also concluded Bayesian prediction limits to be superior in cases with few samples, compared to classic parametric-based prediction limits based solely on the lognormal distribution [20].

Another important feature of a model is how well it can be interpreted by the people who use it for planning purposes. A black-box model provides a certain output (in this case the surgery duration), in a way that makes it hard or impossible to identify the factors that caused the prediction to be this way. So called white-box models allow the user of the pre- diction model to identify which of the factors contributed to the model output and gain an understanding of what causes the model to ’think’ that there will be a delay during the surgery. In terms of stakeholder engagement and trust in the model, a white-box model can be preferable. Regrettably, in the current literature review, many studies do not report which of the factors used in the model turned out to be of significant predictive value, or in other words, which of the factors is important in determining the model outcomes. This makes it harder to identify promising variables to use in predictive models, based on literature research.

4 References

[1] Agarwal, Sheetal et al. “A pervasive computing system for the operating room of the future”. In: Mobile Networks and Applications 12.2-3 (2007), pp. 215–228.
[2] Ahmadi, Seyed-Ahmad et al. “Recovery of surgical workflow without explicit mod- els”. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006. Springer, 2006, pp. 420–428.
[3] Alenizi, Abdullah M et al. “Factors predicting prolonged operative time for individual surgical steps of robot-assisted radical prostatectomy (RARP): A single surgeon’s experience”. In: Canadian Urological Association Journal 9.7-8 (2015), E417.
[4] Ammori, BJ, Larvin, M, and McMahon, MJ. “Elective laparoscopic cholecystectomy: Pre-operative prediction of duration of surgery”. In: Surgical endoscopy 15.3 (2001), pp. 297–300.
[5] Arora, Sonal et al. “The impact of stress on surgical performance: a systematic review of the literature”. In: Surgery 147.3 (2010), pp. 318–330.
[6] Association, Healthcare Financial Management et al. “Achieving operating room efficiency through process integration.” In: Healthcare financial management: journal of the Healthcare Financial Management Associ- ation 57.3 (2003), suppl–1.
[7] Avidan, Michael S et al. “Anesthesia awareness and the bispectral index”. In: New England Journal of Medicine 358.11 (2008), pp. 1097–1108.
[8] Bein, B and Scholz, J. “Supraglottic airway de- vices”. In: Best Practice & Research Clinical Anaesthesiology 19.4 (2005), pp. 581–593.
[9] Bhatia, Beenish et al. “Real-time identification of operating room state from video”. In: AAAI. Vol. 2. 2007, pp. 1761–1766.
[10] Blum, Tobias, Feußner, Hubertus, and Navab, Nassir. “Modeling and segmentation of surgical workflow from laparoscopic video”. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010. Springer, 2010, pp. 400–407.
[11] Blum, Tobias et al. “Workflow mining for visualization and analysis of surgeries”. In: In- ternational journal of computer assisted radi- ology and surgery 3.5 (2008), pp. 379–386.
[12] Bouarfa, Loubna, Jonker, Pieter P, and Dankelman, Jenny. “Discovery of high-level tasks in the operating room”. In: Journal of biomedical informatics 44.3 (2011), pp. 455–462.
[13] Checketts, MR et al. “Recommendations for standards of monitoring during anaesthesia and recovery 2015: Association of Anaesthetists of Great Britain and Ireland”. In: Anaesthesia 71.1 (2016), pp. 85–93.
[14] Combes, Catherine et al. “Using a KDD pro- cess to forecast the duration of surgery”. In: International Journal of Production Eco- nomics 112.1 (2008), pp. 279–293.
[15] Cork, Randall C, Vaughan, Robert W, and Humphrey, Linda S. “Precision and accuracy of intraoperative temperature monitoring.” In: Anesthesia & Analgesia 62.2 (1983), pp. 211–214.
[16] Devi, S Prasanna, Rao, K Suryaprakasa, and Sangeetha, S Sai. “Prediction of surgery times and scheduling of operation theaters in opthol- mology department”. In: Journal of medical systems 36.2 (2012), pp. 415–430.
[17] Dexter, Franklin and Ledolter, Johannes. “Bayesian prediction bounds and comparisons of operating room times even for procedures with few or no historic data”. In: The Journal of the American Society of Anesthesiologists

103.6 (2005), pp. 1259–1167.
[18] Dexter, Franklin et al. “Systematic review of general thoracic surgery articles to identify predictors of operating room case du- rations”. In: Anesthesia & Analgesia 106.4 (2008), pp. 1232–1241.
[19] Dexter, Franklin et al. “Automatic updating of times remaining in surgical cases using bayesian analysis of historical case duration data and “instant messaging” updates from anesthesia providers”. In: Anesthesia & Analgesia 108.3 (2009), pp. 929–940.
[20] Dexter, Franklin et al. “Estimating surgical case durations and making comparisons among facilities: identifying facilities with lower anesthesia professional fees”. In: Anesthesia & Analgesia 116.5 (2013), pp. 1103–1115.
[21] Eger, Edmond I, Saidman, Lawrence J, and Brandstater, Bernard. “Minimum Alveolar Anesthetic Concentration: A Standard of Anesthetic Potency”. In: The Journal of the American Society of Anesthesiologists 26.6 (1965), pp. 756–763.
[22] Eijkemans, Marinus JC et al. “Predicting the unpredictable: A new prediction model for op- erating room times using individual characteristics and the surgeon’s estimate”. In: Anes thesiology 112.1 (2010), pp. 41–9.
[23] Eilers, Helge. “Intravenous anesthetics”. In:
Basics of anesthesia, sixth edition. Ed. By Miller, Ronald D and Pardo, Manuel. Elsevier Health Sciences, 2011. Chap. 9, ”99–114.
[24] Erdogan, S Ayca and Denton, Brian T. “Surgery planning and scheduling”. In: Wi- ley Encyclopedia of Operations Research and Management Science (2011).
[25] Faisal, Aldo. CO424 - Machine Learning and Neural Computation. Imperial College Lon- don, Lecture Notes. 2015.
[26] Forestier, Germain, Riffaud, Laurent, and Jan- nin, Pierre. “Automatic phase prediction from low-level surgical activities”. In: International journal of computer assisted radiology and surgery 10.6 (2015), pp. 833–841.
[27] Franke, Stefan, Meixensberger, Jurgen, and Neumuth, Thomas. “Intervention time prediction from surgical low-level tasks”. In: Jour- nal of Biomedical Informatics 46.1 (2013), pp. 152–159.
[28] Franke, Stefan, Meixensberger, Jurgen, and Neumuth, Thomas. “Multi-perspective work- flow modeling for online surgical situation models”. In: Journal of biomedical informatics 54 (2015), pp. 158–166.
[29] Friedman, Jerome H. “Multivariate adaptive regression splines”. In: The annals of statistics (1991), pp. 1–67.
[30] Gambadauro, Pietro, Campo, Vincenzo, and Campo, Sebastiano. “How Predictable Is the Operative Time of Laparoscopic Surgery for Ovarian Endometrioma?” In: Minimally invasive surgery 2015 (2015).
[31] Ghoneim, MM and Block, RI. “Learning and consciousness during general anesthesia.” In: Anesthesiology 76.2 (1992), p. 279.
[32] Gomes, Carlos et al. “Integrating data mining and optimization techniques on surgery scheduling”. In: Advanced Data Mining and Applications. Springer, 2012, pp. 589–602.
[33] Guedon, ACP et al. “A RFID specific participatory design approach to support design and implementation of real-time location systems in the operating room”. In: Journal of medical systems 39.1 (2015), pp. 1–10.
[34] Guedon, ACP et al. “’It is time to pre- pare the next patient’: Automatic predic- tion of remaining duration in laparoscopic cholecystectomy procedures”. In: (Unpublished manuscript) (2016).
[35] Haji, A et al. “Elective laparoscopic cholecystectomy for surgical trainees: predictive factors of operative time”. In: The Surgeon 7.4 (2009), pp. 207–210.

[36] Hsu, Wen-Chiung et al. “Prediction of operation time for laparoscopic myomectomy by ultrasound measurements”. In: Surgical endoscopy 21.9 (2007), pp. 1600–1606.
[37] Irefi, S and Kopyeva, T. “Perioperative Airway Management”. In: Basic Clinical Anesthesia. Ed. by Sikka, Paul K, Beaman, Shawn T, and Street, James A. Springer, 2015. Chap. 4.
[38] James, Adam et al. “Eye-gaze driven surgical workflow segmentation”. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007. Springer, 2007, pp. 110–117.
[39] Jubran, Amal. “Pulse oximetry”. In: Critical Care 19.1 (2015), pp. 1–7.
[40] Kayis, Enis et al. “Improving prediction of surgery duration using operational and tem- poral factors”. In: AMIA Annual Symposium Proceedings. Vol. 2012. American Medical Informatics Association. 2012, p. 456.
[41] Kayı¸s, Enis et al. “A robust estimation model for surgery durations with temporal, operational, and surgery team effects”. In: Health care management science 18.3 (2015), pp. 222–233.
[42] Kee, Spencer. “Anesthesia and Monitoring Devices”. In: Biomedical Technology and De- vices Handbook. Ed. by Fennigkoh, Larry. CRC Press, 2004. Chap. 22.
[43] Kougias, Panagiotis et al. “Derivation and out- of-sample validation of a modeling system to predict length of surgery”. In: The American Journal of Surgery 204.5 (2012), pp. 563–568.
[44] Kougias, Panagiotis et al. “Modeling anesthetic times. Predictors and implications for short-term outcomes”. In: Journal of Surgical Research 180.1 (2013), pp. 1–7.
[45] Krauss, Baruch and Hess, Dean R. “Capnography for procedural sedation and analgesia in the emergency department”. In: Annals of emergency medicine 50.2 (2007), pp. 172–181.
[46] Lalys, Florent and Jannin, Pierre. “Surgical process modelling: a review”. In: International journal of computer assisted radiology and surgery 9.3 (2014), pp. 495–511.
[47] Lalys, Florent et al. “A framework for the recognition of high-level surgical tasks from video images for cataract surgeries”. In: Biomedical Engineering, IEEE Transactions on 59.4 (2012), pp. 966–976.
[48] Larson, Merlin D. “History of Anesthesia”. In: Basics of anesthesia, sixth edition. Ed. by Miller, Ronald D and Pardo, Manuel. Elsevier Health Sciences, 2011. Chap. 1, ”3–10.
[49] Larsson, Agneta. “The accuracy of surgery time estimations”. In: Production Planning & Control 24.10-11 (2013), pp. 891–902.
[50] Laskin, Daniel M, Abubaker, A Omar, and Strauss, Robert A. “Accuracy of predicting the duration of a surgical operation”. In: Journal of Oral and Maxillofacial Surgery 71.2 (2013), pp. 446–447.
[51] Li, Ying et al. “Predicting surgical case durations using ill-conditioned CPT code matrix”. In: IIE Transactions 42.2 (2009), pp. 121–135.
[52] Loukas, Constantinos and Georgiou, Evangelos. “Surgical workflow analysis with Gaus- sian mixture multivariate autoregressive (GM- MAR) models: a simulation study”. In: Computer Aided Surgery 18.3-4 (2013), pp. 47–62.
[53] Luangkesorn, KL and Eren-Do˘gu, ZF. “Markov Chain Monte Carlo methods for es- timating surgery duration”. In: Journal of Statistical Computation and Simulation 86.2 (2016), pp. 262–278.

[54] Macario, Alex and Dexter, Franklin. “Estimating the duration of a case when the surgeon has not recently scheduled the procedure at the surgical suite”. In: Anesthesia & Analgesia 89.5 (1999), pp. 1241–1245.
[55] Maktabi, Marianne, Vinz, Sascha T, and Neumuth, Thomas. “Frequency based assessment of surgical activities”. In: Current Directions in Biomedical Engineering 1.1 (2015),
pp. 152–156.
[56] Malpani, Anand et al. “System events: readily accessible features for surgical phase detection”. In: International Journal of Computer Assisted Radiology and Surgery (2016), pp. 1– 9.
[57] May, Jerrold H, Strum, David P, and Vargas, Luis G. “Fitting the lognormal distribution to surgical procedure times”. In: Decision Sciences 31.1 (2000), pp. 129–148.
[58] McGrath, Conor D and Hunter, Jennifer M. “Monitoring of neuromuscular block”. In: Con- tinuing Education in Anaesthesia, Critical Care & Pain 6.1 (2006), pp. 7–12.
[59] McKay, Rachel E. “Inhaled Anesthetics”. In: Basics of anesthesia, sixth edition. Ed. by Miller, Ronald D and Pardo, Manuel. Elsevier Health Sciences, 2011. Chap. 8, ”78–98.
[60] Miller, Ronald D. “Scope of Anesthesia Practice”. In: Basics of anesthesia, sixth edition. Ed. by Miller, Ronald D and Pardo, Manuel. Elsevier Health Sciences, 2011. Chap. 2, ”11–19.
[61] Moher, David et al. “Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement”. In: Annals of internal medicine 151.4 (2009), pp. 264–269.
[62] Muller, Meinard. “Dynamic Time Warping”. In: Information retrieval for music and motion. Ed. by Muller, Meinard. Vol. 2. Springer, 2007. Chap. 4.
[63] Nagler, Joshua and Krauss, Baruch. “Capnography: a valuable tool for airway management”. In: Emergency medicine clinics of North America 26.4 (2008), pp. 881–897.
[64] Nakamura, R et al. “Method for end time prediction of brain tumor resections using analysis of surgical navigation information and tumor size characteristics”. In: World Congress on Medical Physics and Biomedical Engineering May 26-31, 2012, Beijing, China. Springer. 2013, pp. 1452–1455.
[65] Padoy, Nicolas et al. “On-line Recognition of Surgical Activity for Monitoring in the Operating Room.” In: AAAI. 2008, pp. 1718–1724.
[66] Padoy, Nicolas et al. “Statistical modeling and recognition of surgical workflow”. In: Medical image analysis 16.3 (2012), pp. 632–641.
[67] Pandit, JJ and Carey, A. “Estimating the duration of common elective operations: impli- cations for operating list management”. In: Anaesthesia 61.8 (2006), pp. 768–776.
[68] Rockstroh, Max et al. “Video-based detection of device interaction in the operating room”. In: Biomedical Engineering/Biomedizinische Technik (2015).
[69] Roth, Patricia. “Anesthesia Delivery Sys- tems”. In: Basics of anesthesia, sixth edition. Ed. by Miller, Ronald D and Pardo, Manuel. Elsevier Health Sciences, 2011. Chap. 15, ”198–219.
[70] Schmidt, Nancy M and Brown, Allan CD. “The Prediction of Surgeon-Specific Operat- ing Room Time (Experience with the Michigan Surgical Monitor)”. In: Proceedings of the Annual Symposium on Computer Application in Medical Care. American Medical Informatics Association. 1984, p. 800.

[71] ShahabiKargar, Zahra et al. "Predicting Pro- cedure Duration to Improve Scheduling of Elective Surgery". In: PRICAI 2014: Trends in Artificial Intelligence. Springer, 2014,
pp. 998–1009.
[72] Shukla, Ramesh K, Ketcham, Jonathan S, and Ozcan, Yasar A. "Comparison of subjective versus data base approaches for improving efficiency of operating room scheduling". In: Health Services Management Research 3.2 (1990), pp. 74–81.
[73] Sikka, Paul K. "History of Anesthesia". In: Basic Clinical Anesthesia. Ed. by Sikka, Paul K, Beaman, Shawn T, and Street, James A. Springer, 2015. Chap. 1.
[74] Silber, Jeffrey H et al. "Influence of patient and hospital characteristics on anesthesia time in medicare patients undergoing general and orthopedic surgery." In: Anesthesiology 106.2 (2007), pp. 356–364.
[75] Silva, Anil da. "Anesthetic Monitoring". In: Basics of anesthesia, sixth edition. Ed. by Miller, Ronald D and Pardo, Manuel. Elsevier Health Sciences, 2011. Chap. 20, "319–331.
[76] Singh, P. M., Shah, D, and Sinha, A. "Anesthesia Machine". In: Basic Clinical Anesthesia. Ed. by Sikka, Paul K, Beaman, Shawn T, and Street, James A. Springer, 2015. Chap. 5.
[77] Spangler, William E et al. "Estimating procedure times for surgeries by determining loca- tion parameters for the lognormal model". In: Health care management science 7.2 (2004), pp. 97–104.
[78] Stauder, R, Okur, A, and Navab, N. "Detecting and Analyzing the Surgical Workflow to Aid Human and Robotic Scrub Nurses". In: The Hamlyn Symposium on Medical Robotics. 2014, p. 91.
[79] Stauder, Ralf et al. "Random forests for phase detection in surgical workflow analysis". In: Information Processing in Computer-Assisted Interventions. Springer, 2014, pp. 148–157.
[80] Stepaniak, Pieter S, Heij, Christiaan, and De Vries, Guus. "Modeling and prediction of surgical procedure times". In: Statistica Neer- landica 64.1 (2010), pp. 1–18.
[81] Stepaniak, Pieter S et al. "Modeling procedure and surgical times for current procedural terminology-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency: a multicenter study". In: Anesthesia & Analge- sia 109.4 (2009), pp. 1232–1245.
[82] Stonemetz, Jerry and Ruskin, Keith. Anesthesia informatics. Springer Science & Business Media, 2008.
[83] Strum, David P, May, Jerrold H, and Var- gas, Luis G. "Modeling the Uncertainty of Sur- gical Procedure Times: Comparison of Log- normal and Normal Models". In: The Journal of the American Society of Anesthesiologists
92.4 (2000), pp. 1160–1167.
[84] Strum, David P et al. "Surgeon and type of anesthesia predict variability in surgical procedure times". In: The Journal of the American Society of Anesthesiologists 92.5 (2000), pp. 1454–1466.
[85] Strum, David P et al. "Estimating Times of Surgeries with Two Component Procedures: Comparison of the Lognormal and Normal Models". In: The Journal of the American So- ciety of Anesthesiologists 98.1 (2003), pp. 232–240.
[86] Strum, DP, May, JH, and Vargas, LG. "Surgical procedure times are well modeled by the lognormal distribution". In: Anesthesia & Analgesia 86.2S (1998), 47S.
[87] Susarla, Srinivas M and Dodson, Thomas
B. "Predicting third molar surgery operative time: a validated model". In: Journal of Oral and Maxillofacial Surgery 71.1 (2013), pp. 5–13.

[88] Tran, Dinh Tuan, Sakurai, Ryuhei, and Lee, Joo-Ho. “An Improvement of Surgical Phase Detection Using Latent Dirichlet Allocation and Hidden Markov Model”. In: Innovation in Medicine and Healthcare 2015. Springer, 2016, pp. 249–261.
[89] Twinanda, Andru P et al. “EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos”. In: arXiv (preprint) (2016).
[90] Vaccaro, Carlos A et al. “Body surface area: a new predictor factor for conversion and prolonged operative time in laparoscopic colorectal surgery”. In: Diseases of the Colon & Rec- tum 55.11 (2012), pp. 1153–1159.
[91] Veen-Berkx, Elizabeth van et al. “The influence of anesthesia-controlled time on operating room scheduling in Dutch university medical centres”. In: Canadian Journal of Anesthesia/Journal canadien d’anesth´esie 61.6 (2014), pp. 524–532.
[92] Wehman, Brody et al. “Patient anatomy predicts operative time in robotic totally endoscopic coronary artery bypass surgery”. In: Interactive cardiovascular and thoracic surgery (2014), ivu226.
[93] Westbury, Sarah, Pandit, Meghana, and Pan- dit, Jaideep J. “Matching surgical operating capacity to demand using estimates of operating times”. In: Journal of health organization and management 23.5 (2009), pp. 554–567.
[94] Wiegmann, Douglas A et al. “Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation”. In: Surgery 142.5 (2007), pp. 658–665.
[95] Wright, Ian H et al. “Statistical Modeling to Predict Elective Surgery Time: Comparison with a Computer Scheduling System and Surgeon-provided Estimates”. In: The Journal of the American Society of Anesthesiologists 85.6 (1996), pp. 1235–1245.
[96] Zhou, Jinshi and Dexter, Franklin. “Method to assist in the scheduling of add-on surgical cases-upper prediction bounds for surgical case durations based on the log-normal distribution”. In: The Journal of the American Soci- ety of Anesthesiologists 89.5 (1998), pp. 1228–1232.
[97] Zhou, Jinshi et al. “Relying solely on historical surgical times to estimate accurately future surgical times is unlikely to reduce the average length of time cases finish late”. In: Journal of clinical anesthesia 11.7 (1999), pp. 601–605.