BUSINESS ANALYTICS PAPER BY MARTINA HOEVER

MONITORING DOOR OPENINGS IN OPERATION ROOMS

Using decision trees to find common factors in operations resulting in high operation room traffic

Supervised by
Rene Bekker

Commissioned by
Kjeld Aij
Frederique Simons

ABSTRACT

Operation room traffic and in particular door openings can lead to distractions and even surgical site infections. From June 2011 until February 2012, 8009 operations in the VU medical center were observed during which 272805 door openings were recorded. Using data mining techniques, a model is built on the data that uses characteristics of operations to find relationships between the different characteristics of an operation and the number of door openings. Since the specialism and the room that the operation took place in seem to have a lot of correlation with the number of door openings, it is advised to start monitoring either per room or per specialism. Monitoring door openings can result in a drop in the average number of door openings in an operation.

SUMMARY

In every operation room several measurements need to be taken to consider its vulnerable environment, for example specific clothing requirements and ventilation. Operation room traffic has a negative effect on this environment. It has been proven that operation room traffic and in particular door openings can lead surgical site infections. It is therefore important to limit the number of door openings in an operation. This study focuses on determining if there are certain characteristics in operations that have a connection with a large number of door openings, so the management of the operation rooms can focus on monitoring operations with these types of characteristics.

The traffic pattern during operations in the VU medical center was observed over the course of six months. Operative details were recorded and a digital counter, manufactured by New Compliance, recorded the door openings. The number of door openings were counted from the first incision until the operation had finished.

The data collected for each operation included the start date, the start time, the end date, the end time, the duration, the room, the specialism, the specialist, the operation code and the operation name. The data collected for each door opening included the date, the time, the sensor ID, the type of door and the room.

Using machine learning techniques, a model was built on the data. This model uses the characteristics of an operation as attributes predict a result. As opposed to 'regular ' machine learning techniques, the goal was not to create this model so it can be used to predict the result for operations that still have to take place. The goal was to use the model to distract which characteristics seem to have the greatest influence on the result. The model uses a decision tree with cross validation, which is calculated in 'RapidMiner'.

From the decision trees that were made, each operation room can be assigned a rank describing whether the number of door openings are very low, low, medium, high or very high. Operation rooms with a very low rank are OK 1, OK 3 and, OK 11. Operation rooms with a low rank are OK 4, OK 6, OK 7 and, OK 13. Operation rooms with a medium rank are OK 5, OK 10 and, OK 12. Operation rooms with a high rank are OK 2, OK 8 and, OK 9. Operation rooms with a very high rank are OK 14, OK 15 and OK 16. The distinction of the different ranks per operating room could be caused by the fact that certain operations usually take place in the same operating room.

By differentiating between the different specialisms and the number of door openings per hour distinctions can be made. Specialisms with very low rank are ANS, OOG and ORT. Specialism with a low rank are NCH, PCH, URO and URK. Specialisms with medium rank are HE and HLO. Specialisms with a high rank are CCH, MON and, TRA. Specialisms with a very high are GYN, HLK, HON, KCH, KNO, ONB, URK and, VER.

By distinguishing the different door types the following conclusions can be drawn. The number of door openings of the setting room and the scrub room seem mostly dependent on the duration of the operation. The total number of openings of the patient entrance is very low compared to the number of openings of the doors to the setting room and the scrub room. It is shown that only for room OK-1 it is predicted that the door to the patient entrance will open. The number of openings of the patient accompanist entrance is also relatively low. It is shown that only the operation code seems to have predictive value for the number of openings of the patient accompanist entrance.

SAMENVATTING

Er worden veel maatregelingen genomen om de kwetsbare omgeving van een operatiekamer beschermd te houden. Deuropeningen hebben een negatief effect op deze omgeving. Het is zelfs bewezen dat een hoog aantal deuropeningen infecties kunnen veroorzaken. Daarom is het belangrijk het aantal deuropeningen zo laag mogelijk te houden. Het doel van dit onderzoek is overeenkomsten te vinden tussen bepaalde eigenschappen van een operatie en een hoog aantal deur openingen. Wanneer deze eigenschappen bekend zijn kan het management van de operatiekamers zich focussen op het monitoren van operaties met deze eigenschappen.

Voor 6 maanden is het loopgedrag gedurende operaties in het VU medisch centrum bijgehouden. Details van alle operaties zijn opgenomen en een digitale teller, geproduceerd door NewCompliance, registreerde de deuropeningen. Het aantal deuropeningen werd geteld vanaf het moment van eerste incisie tot sluiting.

De data die bij elke operatie werd geregistreerd bestond uit de startdatum, de starttijd, de einddatum, de eindtijd, de operatieduur, de kamer, het specialisme, de specialist, de operatiecode en de operatienaam. De data geregistreerd bij elke deuropening bestond uit de datum, het tijdstip, de sensor ID, het type deur en de kamer.

Gebruik makend van ‘Machine learning techniques’ is een model gemaakt om de data te analyseren. Dit model gebruikt de eigenschappen van de operaties als attributen om een resultaat te voorspellen. In tegenstelling tot ‘gewone’ Machine learning techniques is het doel niet om dit model te gebruiken om het resultaat voor operaties in de toekomst te voorspellen, maar uit te vinden welke eigenschappen van een operatie geselecteerd worden om het resultaat te voorspellen. Het model gebruikt een beslissingsboom met cross validation, dat bepaald wordt met behulp van het programma RapidMiner.

Met behulp van de gemaakte beslissingsbomen kan aan elke operatie kamer een rank worden toegewezen, welke beschrijft of het aantal deuropeningen heel laag, laag, gemiddeld, hoog of heel hoog zal zijn. Operatiekamers met een heel laag aantal deur openingen zijn OK 1, OK 3 en OK 11. Operatiekamers met een laag aantal deur openingen zijn OK 4, OK 6, OK 7 en OK 13. Operatiekamers met een gemiddeld aantal deur openingen zijn OK 5, OK 10 en OK 12. Operatiekamers met een hoog aantal deur openingen zijn OK 2, OK 8 en OK 9. Operatiekamers met een heel hoog aantal deur openingen zijn OK 14, OK 15 en OK 16.

Aan de verschillende specialismes kan ook een rank worden toegewezen. Specialismes met een heel laag aantal deuropeningen zijn ANS, OOG en ORT. Specialismes met een laag aantal deuropeningen zijn NCH, PCH, URO en URK. Specialismes met een gemiddeld aantal deuropeningen zijn HE en HLO. Specialismes met een hoog aantal deuropeningen zijn CCH, MON en TRA. Specialismes met een heel hoog aantal deuropeningen zijn GYN, HLK, HON, KCH, KNO, ONB, URK en VER.

Wanneer onderscheid wordt gemaakt tussen de verschillende deurtypes kunnen de volgende conclusies getrokken worden. Het aantal deuropeningen van de opdekkamer en de wasruimte blijkt vooral afhankelijk van de duur van de operatie. Het aantal deuropeningen van de patiënten ingang is erg laag vergeleken met het aantal deuropeningen van de opdekkamer en de wasruimte. Alleen bij OK-1 werd voorspeld dat deze deur open zou gaan. Het blijkt dat de operatie code als enige een voorspellende waarde heeft voor het openen van de patiënten begeleiders ingang.

PREFACE

This paper was written as an assignment for my Master in Business Mathematics & Informatics. The goal of the assignment is to write a paper on a research subject combining the business, mathematics and computer science aspects of the Master.

As a working-student in the VU medical center, I wanted to perform my research on a subject that would be of value to the hospital.

I would like to thank:

Kjeld Aij

Frederique Simons

René Bekker

TABLE OF CONTENTS

1 INTRODUCTION

In every operation room several measurements need to be taken to consider its vulnerable environment, for example specific clothing requirements and ventilation. Operation room traffic has a negative effect on this environment. It has been proven that operation room traffic and in particular door openings can lead to distractions and even surgical site infections.[1] Some movement within a theatre complex is required for clinical reasons. Other movements may be logistically orientated, including change of theatre personnel, staff comfort breaks, movement of equipment from communal storage areas and communication between staff. Some traffic cannot be anticipated and there is always the potential for an emergency situation requiring specialist staff or equipment from outside the theatre.

However, even considering these inevitable door openings, the number of door openings in the average operation seems unnecessary high. Meanwhile, it has been shown that door openings disturb the engineered airflow of the operation room, which can seed the air above the incision with bacteria from the patient's skin, the operative staff, and areas outside the OR.[2] Surgical-site infections are still a major problem in modern medicine. These infections may be deep or superficial. The superficial ones are easier to deal with, but the deep infections can be complicated, leading to re-operation, or can even be life threatening. In addition, researchers have explored the potential factors which contribute to mistakes during surgical operations which showed that distractions including door opening and increased theatre traffic have been identified as contributory factors to surgical mistakes.[3]

It is therefore important to limit the number of door openings in an operation. This study focuses on determining if there are certain characteristics in operations that have a connection with a large number of door openings, so the management of the operation rooms can focus on monitoring operations with these types of characteristics.

This report will describe how a model is built using data mining techniques on a Data set acquired from the recordings in operation rooms in the VU medical center. After exploring the data, different smaller Data sets are created. Using machine learning techniques, a model is built on these Data sets. This model uses the characteristics of an operation as attributes to predict a result as for instance 'the number of door openings' or 'the number of door openings per hour'. As opposed to 'regular ' machine learning techniques, the goal is not to create this model so it can be used to predict the result for operations that still have to take place. The goal is to use the model to distract which characteristics seem to have the greatest influence on the result. The model uses a decision tree with cross validation. If these characteristics can be identified, monitoring operations with these types of characteristics might result in a lower number of door openings.

2 MATERIALS AND METHODS

The traffic pattern during operations in the VU medical center was observed over the course of six months. Operative details were recorded and a digital counter, manufactured by New Compliance, recorded the door openings. The number of door openings were counted from the first incision until the operation had finished.

2.1 THE DATA

New Compliance provided two separate .csv files, one that contains all the surgeries and one that contains all the door openings that took place between June 2011 and February 2012. The initial Data set consists of the following data:

The data collected for each operation included:

- Start Date
- Start Time
- End Date
- End Time
- Duration
- Room
- Specialism
- Specialist
- CTCG (Operation code)
- CRGD (Operation name)

The data collected for each door opening included:

- Date
- Time
- Sensor ID
- Type of door
- Room

Type of door can either be setting room, scrub room, patient entrance or patient accompanist entrance. This distinction is made since the different doors are different sizes and lead to different areas. It is assumed that these distinctions lead to different changes in the air flow. The necessity of door openings also differs per door type. There are, for instance, not a lot of reasons the patient entrance or the patient accompanist entrance should be opened, since the counting of door openings starts after the patient is in the operating room.

A total of 9089 operations and 1052250 door openings were registered. There were 11 operations with a negative duration, these were considered measurement errors. Furthermore operations with a duration below 10 minutes were omitted from this study since they are not representative for this study. These alterations resulted in a Data set with 8009 operations where 272805 door openings were recorded.

By assigning the door openings to the specific operations, using a program written in Java, a data model was created, of which Figure 1 is a graphical representation.

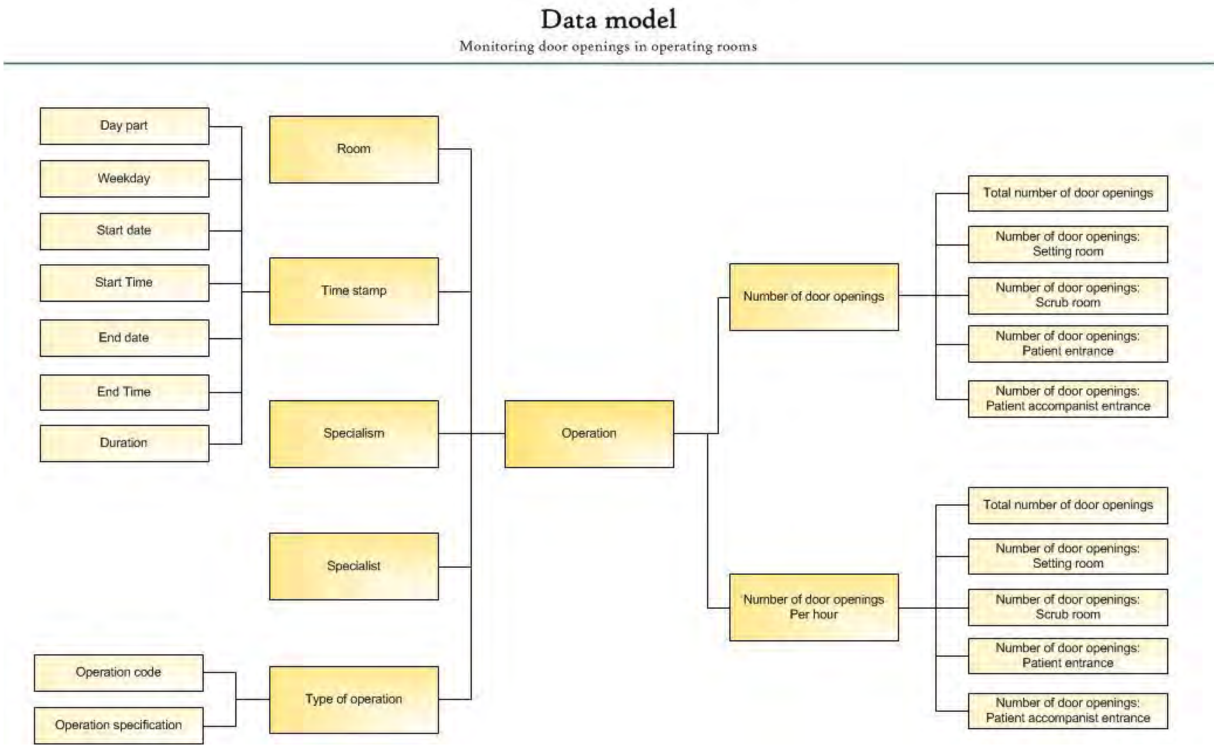

FIGURE 1: DATA MODEL

In this model 'Day part' describes whether the operation started in the morning (06:00 - 12:00), midday (12:00-18:00), evening (18:00 - 00:00) or night (00:00 - 06:00) and 'Weekday' describes the day of the week that the surgery took place.

The number of door openings per surgery ranged from 0 to 555. In 230 operations 0 door openings took place and the average number of door openings is 32. The frequency of door openings is shown in the histogram in Figure 2.

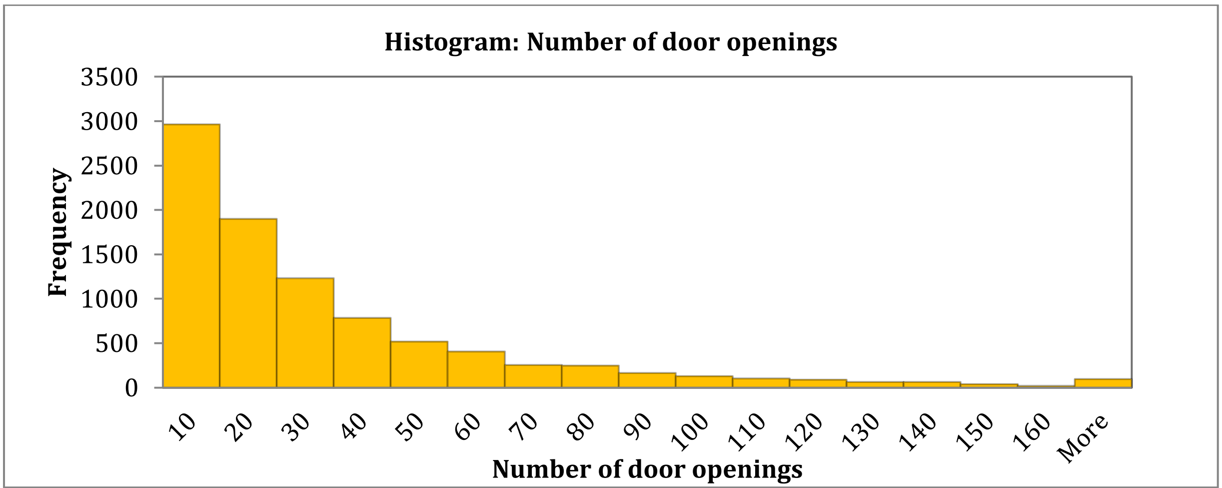

FIGURE 2: HISTOGRAM OF NUMBER OF DOOR OPENINGS

It can be expected that the longer the operation, the higher number of door openings. To investigate the relationship between these variables, they are plotted in Figure 3. It shows that a linear relation between the variables is very likely.

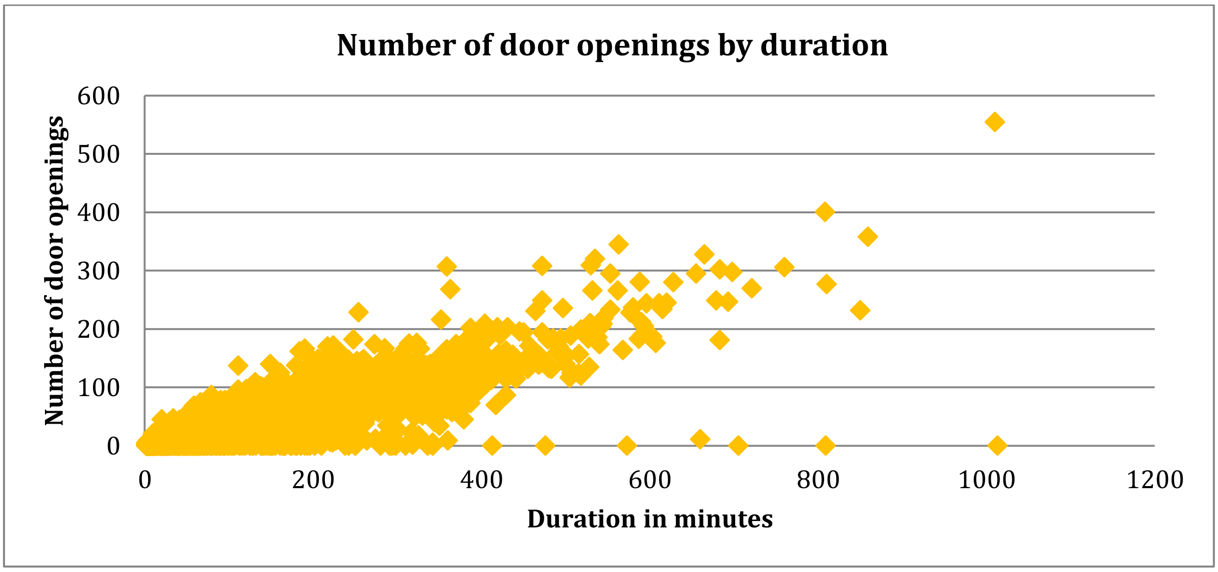

FIGURE 3: NUMBER OF DOOR OPENINGS BY DURATION

Figure 4 shows how the 8009 door openings were distributed over the different doors. As expected, the patient entrance and patient accompanist entrance together only concerned 5% of the door openings.

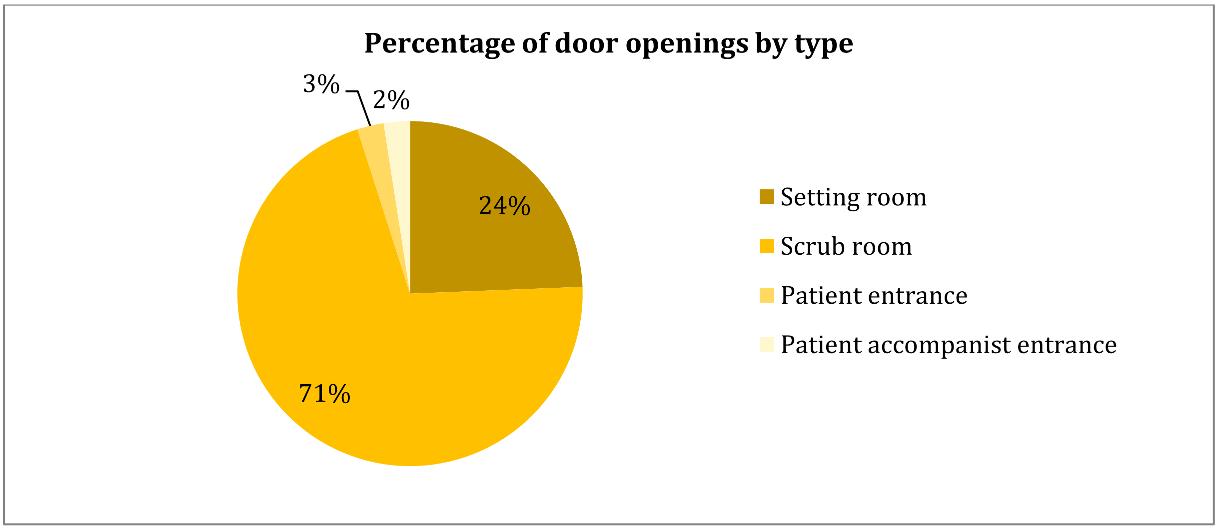

FIGURE 4: PERCENTAGE OF DOOR OPENINGS BY TYPE

2.1.1 CORRELATION BETWEEN THE CHARACTERISTICS

To get an initial idea about the relationships between the different variables, the correlation coefficients are calculated. A correlation coefficient gives an idea if two variables have an interdependence between them. If a correlation coefficient is ± 1 there is a straight line relationship between the variables. If a correlation coefficient is 0 then the variables are said to be uncorrelated.[4]

An example of correlated variables:

The number of people that bring an umbrella on a particular day is positively correlated with the amount of rain that falls on that particular day. The correlation coefficient between these variables can be assumed to be between 0 and 1.

For two quantitative variables, for instance start time and number of door openings, a correlation coefficient can be calculated using the Pearson correlation coefficient. A correlation coefficient between two quantitative variables X and Y with expected values μ_x and μ_y and standard deviations σ_x and σ_y is defined by:

$$\rho_{X,Y} = \frac{E\left[(X-\mu_x)(Y-\mu_y)\right]}{\sigma_x \sigma_y}$$

The correlation coefficients between the characteristics, the number of door openings and the number of door openings per hour are shown in Table 1. The darker the colour, the higher the positive or negative correlation.

TABLE 1: CORRELATION COEFFICIENTS BETWEEN CHARACTERISTICS

	Day part	Weekday	Start date	Start time	End date	End time	Duration	Room	Specialism	Specialist	Operation Code	Operation name	Operation name short	Number of door openings	Number of door openings ph
Day part	1.00	0.00	0.00	0.66	0.00	0.46	-0.18	0.05	0.00	-0.03	0.00	-0.07	-0.05	-0.16	0.03
Weekday	0.00	1.00	-0.02	-0.02	-0.02	-0.01	0.01	-0.02	0.00	-0.01	0.01	0.01	0.05	0.00	-0.01
Start date	0.00	-0.02	1.00	0.01	1.00	0.02	0.04	0.01	-0.04	0.11	0.04	0.08	0.06	0.10	0.16
Start time	0.66	-0.02	0.01	1.00	0.01	0.77	-0.16	0.03	-0.04	-0.02	0.02	-0.06	-0.03	-0.15	0.01
End date	0.00	-0.02	1.00	0.01	1.00	0.02	0.04	0.01	-0.04	0.11	0.04	0.08	0.06	0.10	0.16
End time	0.46	-0.01	0.02	0.77	0.02	1.00	0.25	0.10	0.04	0.03	0.01	0.14	0.05	0.28	0.02
Duration	-0.18	0.01	0.04	-0.16	0.04	0.25	1.00	0.21	0.19	0.06	-0.02	0.36	0.15	0.86	-0.17
Room	0.05	-0.02	0.01	0.03	0.01	0.10	0.21	1.00	0.34	-0.05	-0.08	0.12	0.02	0.20	-0.04
Specialism	0.00	0.00	-0.04	-0.04	-0.04	0.04	0.19	0.34	1.00	0.08	-0.04	0.10	0.02	0.19	-0.01
Specialist	-0.03	-0.01	0.11	-0.02	0.11	0.03	0.06	-0.05	0.08	1.00	0.05	0.14	0.15	0.08	0.09
Operation Code	0.00	0.01	0.04	0.02	0.04	0.01	-0.02	-0.08	-0.04	0.05	1.00	0.42	0.28	-0.02	0.00
Operation name	-0.07	0.01	0.08	-0.06	0.08	0.14	0.36	0.12	0.10	0.14	0.42	1.00	0.49	0.38	0.07
Operation name short	-0.05	0.05	0.06	-0.03	0.06	0.05	0.15	0.02	0.02	0.15	0.28	0.49	1.00	0.16	0.05
Number of door openings	-0.16	0.00	0.10	-0.15	0.10	0.28	0.86	0.20	0.19	0.08	-0.02	0.38	0.16	1.00	0.16
Number of door openings ph	0.03	-0.01	0.16	0.01	0.16	0.02	-0.17	-0.04	-0.01	0.09	0.00	0.07	0.05	0.16	1.00

The bottom two rows in Table 1 show the correlation coefficients between the different characteristics and number of door openings (per hour). As expected from Figure 3, a high correlation exists between the duration of an operation and the number of door openings. A correlation coefficient is considered reasonably if it is above 0.10.

Reasonably high correlation with number of door openings:

- Day part
- Start date
- Start time
- End time
- End date
- Duration
- Room
- Specialism
- Operation name short

Reasonably high correlation with number of door openings per hour:

- Start date
- End date
- Duration

Using this information we can build a prediction model that uses a data set to predict a number of door openings for an operation with certain characteristics.

2.2 THE MODEL

Using machine learning techniques, a model is built on the data. This model uses the characteristics of an operation as attributes predict a result. As opposed to 'regular ' machine learning techniques, the goal is not to create this model so it can be used to predict the result for operations that still have to take place. The goal is to use the model to distract which characteristics seem to have the greatest influence on the result. The model uses a decision tree with cross validation, which is calculated in 'RapidMiner'.

2.2.1 DECISION TREE

Decision tree learning is a machine learning technique that uses a decision tree as a predictive model that predicts the value of a target variable - the result - based on several input values - the attributes -. In this study classification trees are used; a decision tree that describes data but not decisions; rather, the resulting classification tree can be an input for decision making. Classification trees are used to predict membership of objects in the classes of a categorical dependent variable from their measurements on one or more predictor variables. In this case the classes are represented by the different ranks.

A tree works by splitting the Data set into subsets based on an attribute value test. First, select an attribute to place at the root node, and make one branch for each possible value. This splits up the Data set in two subsets, one for every value of the attribute. Now, the process can be repeated recursively for each branch, using only those instances that actually reach the branch. The recursion is completed when the subset at a node are all of the same result, or when splitting no longer adds value to the predictions. To determine which attribute to split on at each point in the tree, the gain ratio is calculated for each attribute. The gain ratio is derived by taking into account the number and size of daughter nodes into which an attribute splits the Data set, disregarding any information of the class.

An example of a classification tree:

Imagine someone wants to decide whether or not to take an umbrella to work. From past experience they know that if it is not sunny, an umbrella is needed. They also know that when it is sunny they only need an umbrella if it is cloudy. From this information a decision tree can be created displayed in Figure 1.

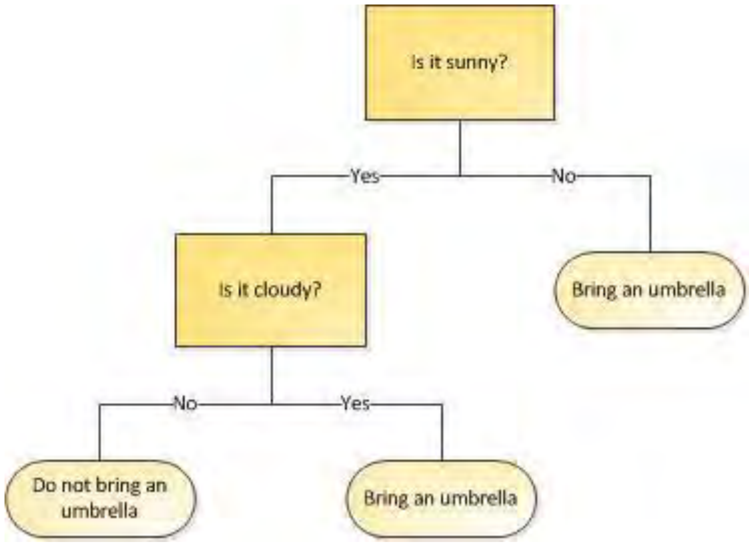

FIGURE 5: DECISION TREE EXAMPLE

This classification tree can be created from a Data set where there is registered for a certain amount of days whether it was sunny, cloudy and if an umbrella was needed. If this data implicates that almost every time it was sunny an umbrella was not needed, and if it was not sunny an umbrella was only needed if it also was cloudy, this decision tree would be the result of data mining.

2.2.2 CROSS VALIDATION

Cross-validation is a technique for assessing how the results of a statistical analysis will generalize to an independent Data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. A part of the available Data set is not used to create the classification tree, but kept separately as a test set. When the tree is finished, the test set is used to test whether or not the tree would predict the correct classification.

Cross validation applied to the previous example:

If in the test set 20% of the time an umbrella was not needed when it was not sunny, it would name these predictions as inaccurate. If all the other predictions were accurate, the model would have an accuracy of 80%.

2.3 THE DATA SETS

From the recorded data different smaller Data sets are created to determine the relationships between the different variables with the prediction model. The different data sets have either a different subset of the characteristics, or a different result, like 'total number of door openings' and 'total number of door openings per hour'. The different subsets of characteristics are chosen to explore if different combinations will result in the selection of different attributes.

Since the range of number of door openings per operation is very large, from a minimum of zero door openings to a maximum of 500 door openings, it does not seem very useful to predict the exact number of door openings for each operation. The difference between 450 and 451 door openings does not seem to matter that much. Therefore, each data set is assigned a rank based on the percentile by sorting the number of door openings. The names of the ranks and the associated percentile are displayed in Table 2.

TABLE 2: RANKS PER PERCENTILE

Rank	Percentile
Very Low	0% - 20%
Low	20% - 40%
Medium	40% - 60%
High	60% - 80%
Very High	80% - 100%

2.3.1 DATA SET 1

In data set 1 all attributes are used to predict the rank 'total number of door opening'. According to the distribution of the number of door openings, a rank was assigned to every operation. Sorting the operations on number of door openings, the operations are split into 5 percentiles. The ranks, that represent in which percentile the operation falls, are displayed in Table 3.

TABLE 3: RANK OF OPERATIONS BY NUMBER OF DOOR OPENINGS

Rank	Number of door openings	Percentile
Very Low	[0, 7]	0% - 20%
Low	[8, 15]	20% - 40%
Medium	[16, 27]	40% - 60%
High	[28, 50]	60% - 80%
Very High	[50, 550]	80% - 100%

2.3.2 DATA SET 2

In data set 2 all attributes used to predict the rank 'total number of door openings per hour'. Since a linear relationship between duration and number of door openings is expected it can lead to interesting results if the number of door openings is converted to number of door openings per hour. The ranks, that represent in which percentile the operation falls, are displayed in Table 4.

TABLE 4: RANK OF OPERATIONS BY NUMBER OF DOOR OPENINGS PER HOUR

Rank	Number of door openings	Percentile
Very Low	[0, 14.14]	0% - 20%
Low	[14.14, 19.59]	20% - 40%
Medium	[19.59, 24.55]	40% - 60%
High	[24.55, 32.31]	60% - 80%
Very High	[32.31, 136]	80% - 100%

2.3.3 DATA SET 3

In data set 3 the rank 'total number of door openings' is predicted by the attributes that show a correlation with the total number of door openings. These are the following variables:

- Day part
- Start date
- Start time
- End time
- End date
- Duration
- Room
- Specialism
- Operation name short

2.3.4 DATA SET 4

In data set 4 the rank 'total number of door openings per hour' is predicted by the attributes that show a correlation with the total number of door openings per hour. These are the following variables:

- Start date
- End date
- Duration

2.3.5 DATA SET 5

Since the different doors open for different reasons and have different consequences, Data set 5 uses all attributes to predict the total number of door openings per door type.

2.3.5.1 DATA SET 5A

Data set 5A predicts the total number of setting room door openings. The ranks, that represent in which percentile the operation falls, are displayed in Table 5.

TABLE 5: RANK OF OPERATIONS BY NUMBER OF SETTING ROOM DOOR OPENINGS

Rank	Number of door openings	Percentile
Very Low	[0, 1]	0% - 20%
Low	[1, 4]	20% - 40%
Medium	[4, 7]	40% - 60%
High	[7, 12]	60% - 80%
Very High	[12, 157]	80% - 100%

2.3.5.2 DATA SET 5B

Data set 5B predicts the total number of scrub room door openings. The ranks, that represent in which percentile the operation falls, are displayed in Table 6.

TABLE 6: RANK OF OPERATIONS BY NUMBER OF SCRUB ROOM DOOR OPENINGS

Rank	Number of door openings	Percentile
Very Low	[0, 6]	0% - 20%
Low	[6, 12]	20% - 40%
Medium	[12, 21]	40% - 60%
High	[21, 37]	60% - 80%
Very High	[37, 400]	80% - 100%

2.3.5.3 DATA SET 5C

Data set 5C predicts the total number of times the patient entrance opens. Since the patient entrance only represents 3% of all door openings, the rank for each operation represents if the patient entrance has opened at least once in an operation. This is displayed in Table 7.

TABLE 7: RANK OF OPERATIONS BY NUMBER OF PATIENT ENTRANCE OPENINGS

Rank	Number of door openings	Percentile
Low	0	0% - 76%
High	>0	76% - 100%

2.3.5.4 DATA SET 5D

Data set 5D predicts the total number of times the patient accompanist entrance opens. Since the patient entrance only represents 2% of all door openings, the rank for each operation represents if the patient accompanist entrance has opened at least once in an operation.

TABLE 8: RANK OF OPERATIONS BY NUMBER OF PATIENT ACCOMPANIST ENTRANCE OPENINGS

Rank	Number of door openings	Percentile
Low	0	0% - 95%
High	>0	95% - 100%

2.3.6 DATA SET 6

In data set 6 the attribute specialism is used to predict the rank 'total number of door openings per hour'. To provide a representative result only specialisms that had at least 100 operations in the Data set are considered. The ranks, that represent in which percentile the operation falls, are displayed in Table 4.

3 RESULTS

For each different data set a decision tree is created. The decision trees show how the data sets can be divided in the different classes, displayed by their ranks. For each decision tree the results from cross validation are shown to describe how accurate the results are. This shows what percentage of the test data resulted in an accurate prediction, what percentage was predicted one rank off, and what percentage was predicted more than one rank off.

3.1 DECISION TREES

3.1.1 DATA SET 1

Data set 1 consists of all the attributes and predicts the rank 'number of door openings'. As shown in Figure 6 the decision tree for data set 1 is largely based on duration. Only if an operation is longer than 142.5 minutes the attribute day part seems to have an influence. It seems that in that case, operations that take place midday and in the morning have more door openings than operations that take place at night.

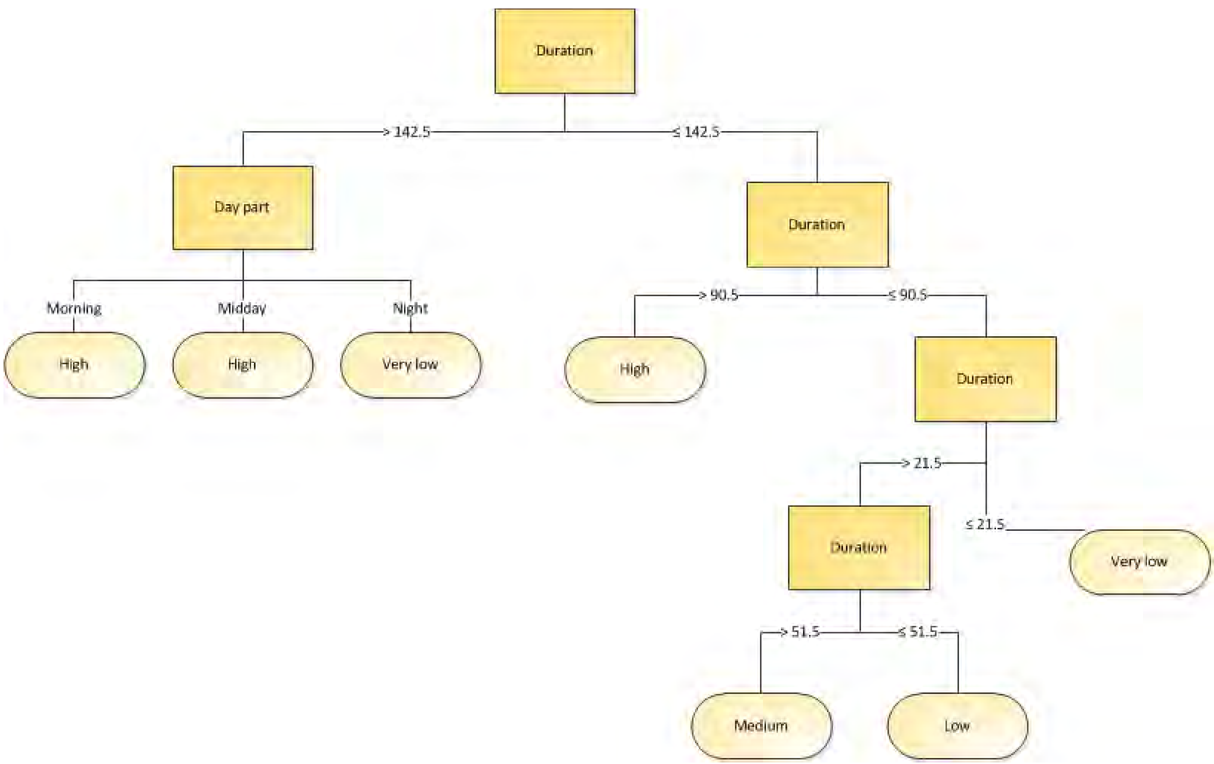

FIGURE 6: DECISION TREE FOR DATA SET 1

3.1.2 DATA SET 2

Data set 2 consists of all the attributes and predicts the rank 'number of door openings per hour'. As shown in Figure 7 the decision tree for data set 2 is completely based on the attribute room.

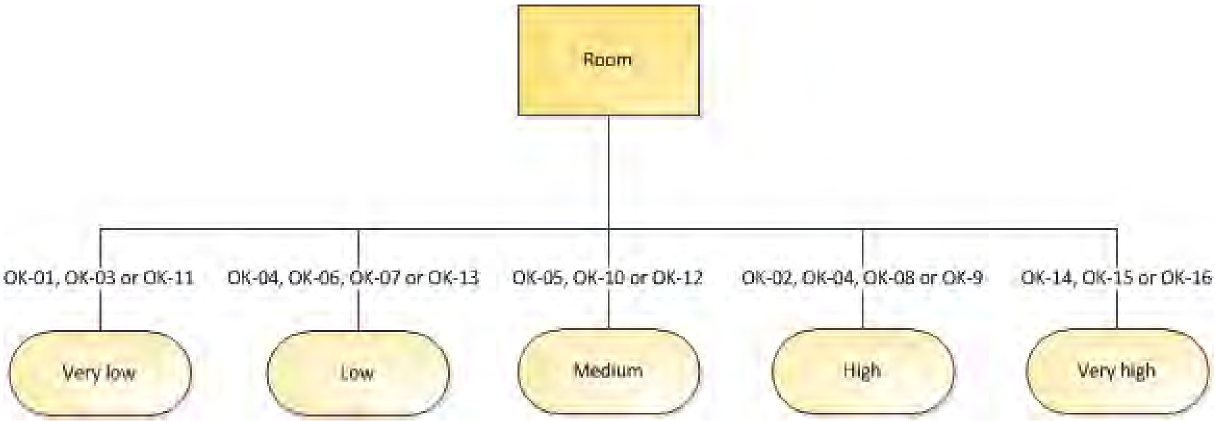

FIGURE 7: DECISION TREE FOR DATA SET 2

From this decision tree the operating rooms can be sorted by their rank 'number of door openings per hour':

Operation rooms with a very low rank:

- OK 1
- OK 3
- OK 11

Operation rooms with a low rank:

- OK 4
- OK 6
- OK 7
- OK 13

Operation rooms with a medium rank:

- OK 5
- OK 10
- OK 12

Operation rooms with a high rank:

- OK 2
- OK 8
- OK 9

Operation rooms with a very high rank:

- OK 14
- OK 15
- OK 16

3.1.3 DATA SET 3

Data set 3 consists of the attributes with a reasonably high correlation with the number of door openings and predicts the rank 'number of door openings'. The decision tree for Data set 3 displayed Figure 8 is, like the decision tree for the similar Data set 1, largely based on duration. However this decision tree also shows that start date, end time and operation room also have some influence.

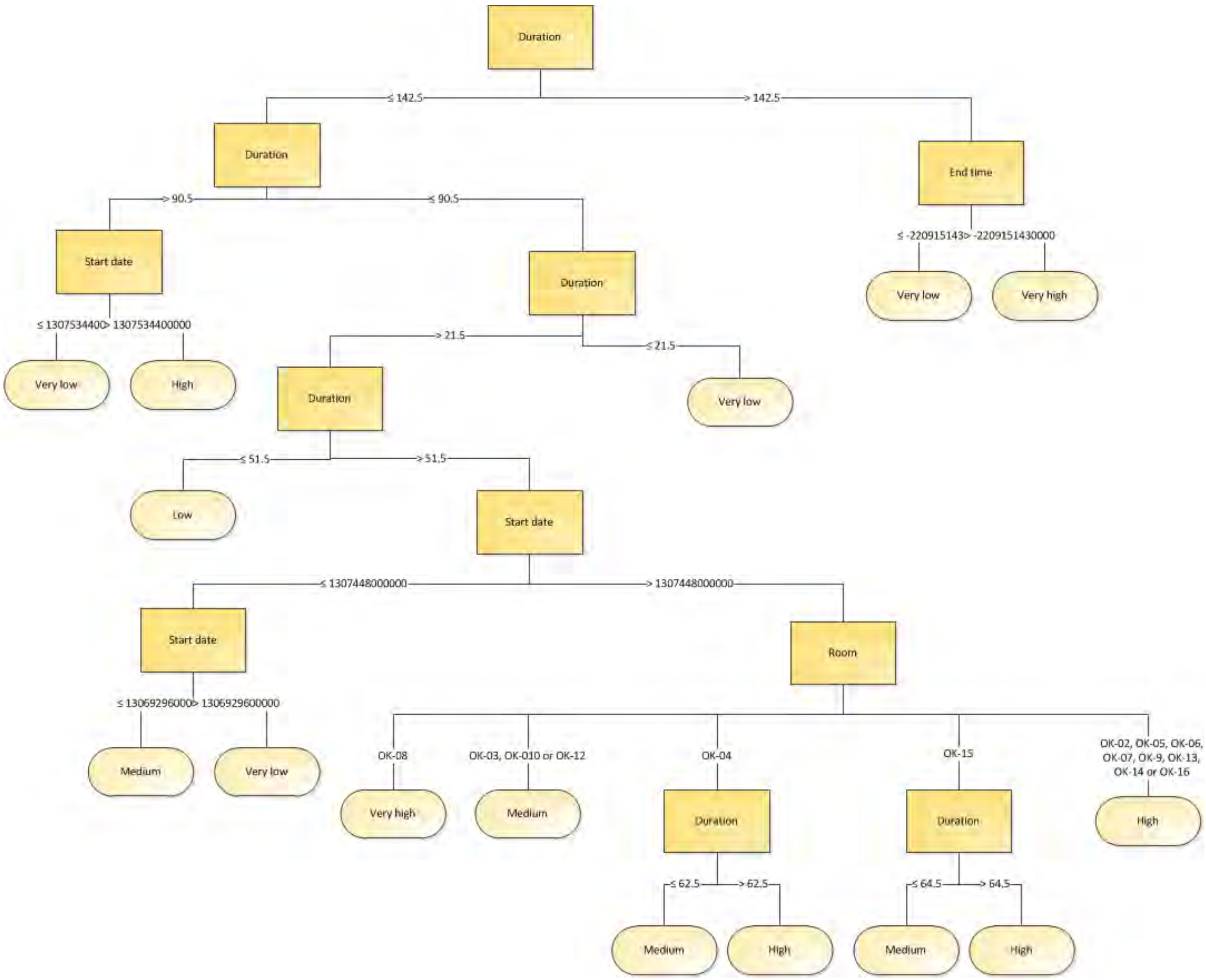

FIGURE 8: DECISION TREE FOR DATA SET 3

3.1.4 DATA SET 4

Data set 4 consists of the attributes with a reasonably high correlation with the number of door openings and predicts the rank 'number of door openings'. The decision tree based on Data set 4, displayed in Figure 9, shows that the attributes start date end date and duration don't have enough predictive value to build an accurate model.

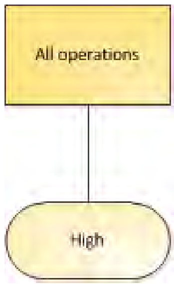

FIGURE 9: DECISION TREE FOR DATA SET 4

3.1.5 DATA SET 5A

Data set 5A consists of all the attributes and predicts the rank ‘number of door openings: setting room’. The decision tree for Data set 5A, displayed in Figure 10, shows that the following attributes can predict the number of door openings of the setting room:

- Duration
- Start date
- End time
- Room

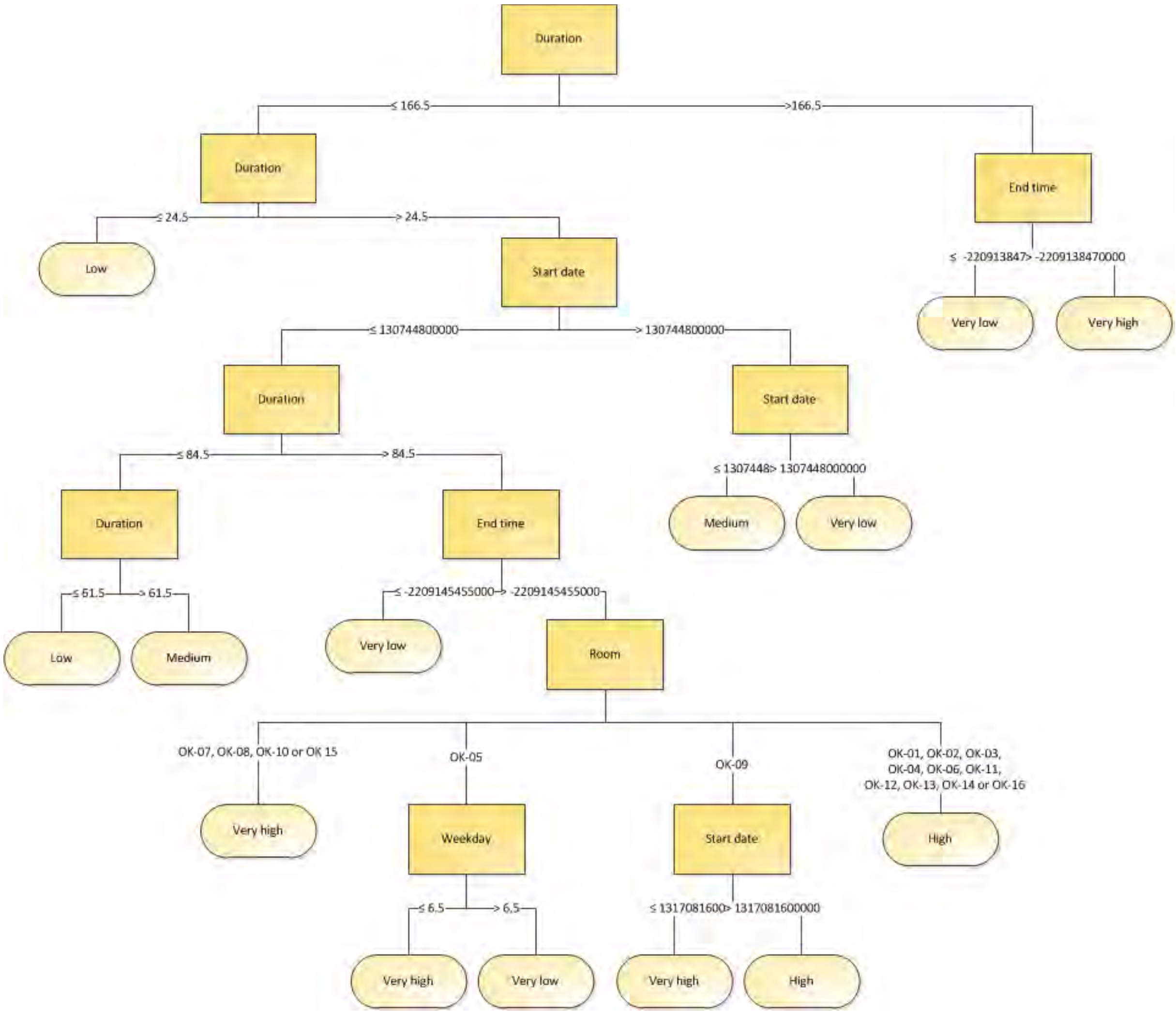

FIGURE 10: DECISION TREE FOR DATA SET 5A

3.1.6 DATA SET 5B

Data set 5B consists of all the attributes and predicts the rank 'number of door openings: scrub room'. The decision tree for data set 5B, displayed in Figure 11, shows that the following attributes can predict the number of door openings of the scrub room:

- Duration
- Start date
- Room

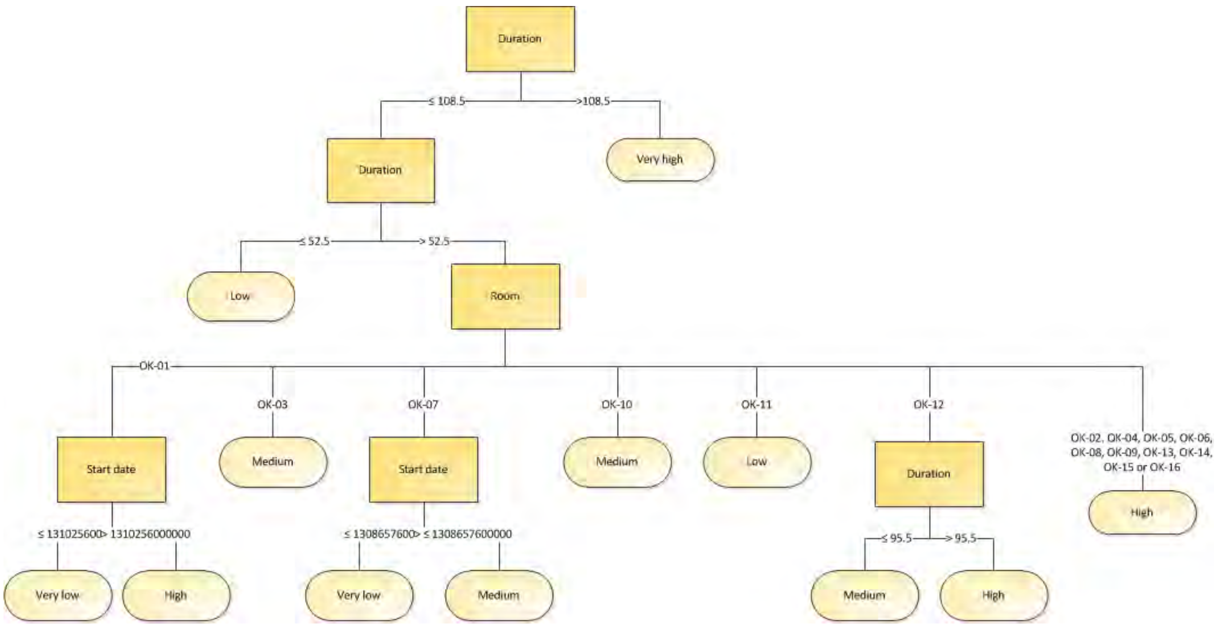

FIGURE 11: DECISION TREE FOR DATA SET 5B

3.1.7 DATA SET 5C

Data set 5C consists of all the attributes and predicts the rank 'number of door openings: patient entrance'. The decision tree for data set 5C, displayed in Figure 12, shows that only for room OK-1 it is predicted that the door to the patient entrance will open. For OK-1 this seems to depend on the following attributes:

- Duration
- Start date
- Room

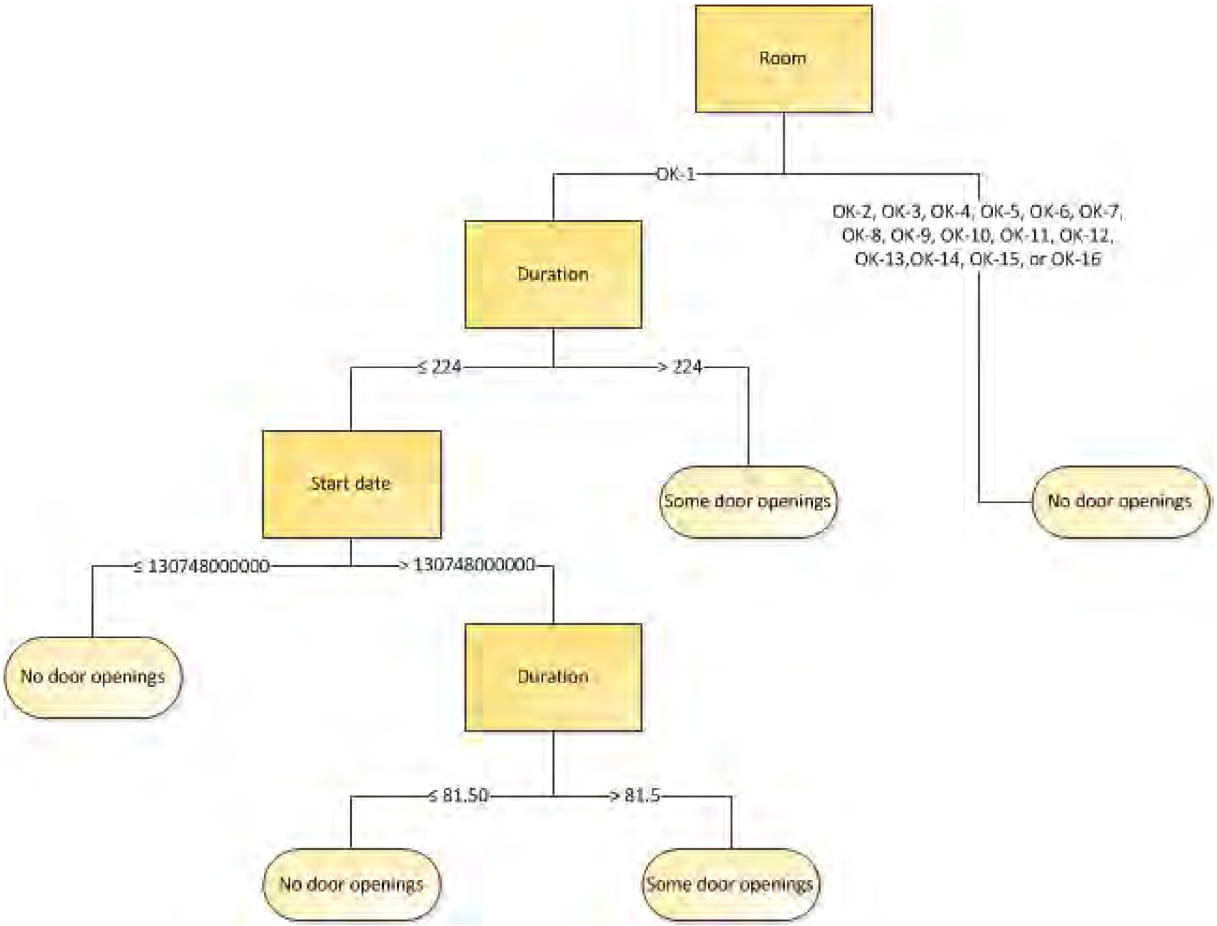

FIGURE 12: DECISION TREE FOR DATA SET 5C

3.1.8 DATA SET 5D

Data set 5D consists of all the attributes and predicts the rank ‘number of door openings: patient accompanist entrance’. The decision tree for Data set 5D, displayed in Figure 13, shows that only the operation code has a predictive value for the number of openings of the patient accompanist entrance:

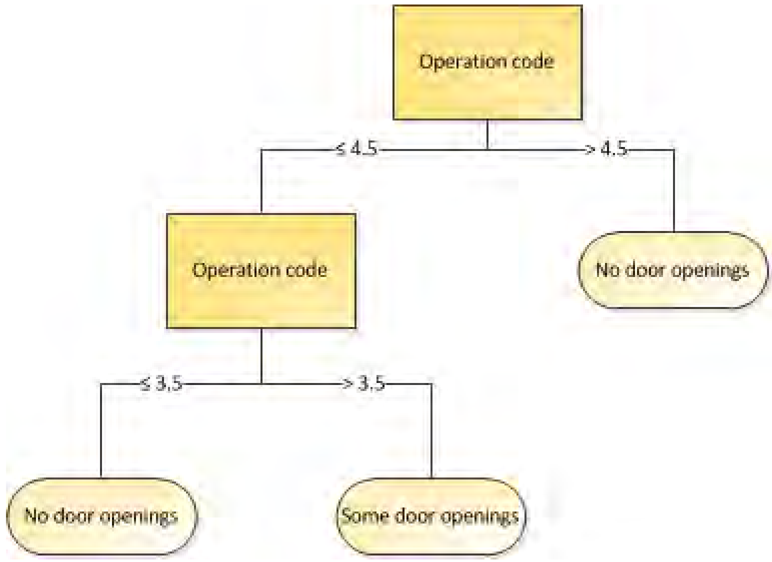

FIGURE 13: DECISION TREE FOR DATA SET 5D

3.1.9 DATA SET 6

Data set 6 consists of the attribute specialism and predicts the rank 'number of door openings per hour'. From the decision tree for Data set 6, shown in Figure 14, the different specialisms can be sorted by their rank: total number of door openings per hour.

Specialisms with a very low rank:

- ANS
- OOG
- ORT

Specialism with a low rank:

- NCH
- PCH
- URO
- URK

Specialisms with a medium rank:

- HGE
- HLO

Specialisms with a high rank:

- CCH
- MON
- TRA

Specialisms with a very high rank:

- GYN
- HLK
- HON
- KCH
- KNO
- ONB
- URK
- VER

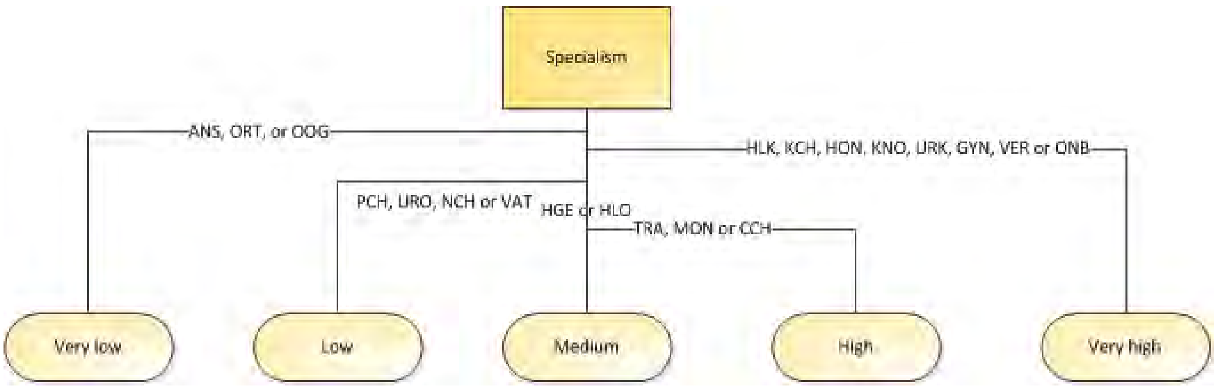

FIGURE 14: DECISION TREE FOR DATA SET 6

3.2 CROSS VALIDATION

When making the decision trees, part of the data was kept separate as a test set. This test set is used to determine the accuracy of the predictions using cross validation.

The results of cross validation on the decision tree for data set 1 show that 53% of the predictions was accurate and only 9% of the predictions was more than one rank off. The results of cross validation on the decision tree for data set 3 show that the percentage of correct predictions is slightly higher than that for data set 1. Furthermore the percentage of prediction that were more than one rank off is slightly lower than that for data set 1. It seems that data set 3 shows a slightly more accurate prediction.

For data set 2, 30% of the predictions was accurate and 37% of the predictions was more than one rank off. As expected from the decision tree for data set 4, cross validation shows that only 20% of the predictions are accurate. Even 60% of the predictions are more than one rank off. It shows that data set 2 shows a more accurate prediction than data set 4.

Cross validation showed that the data sets 5A, 5B, 5C, and 5D are reasonably accurate with 23%, 11%, 24% and 5% of the preditions more than one rank off.

The results of cross validation on data set 6 show that 30% of the predictions were accurate and 39% of the predictions are more than one rank off.

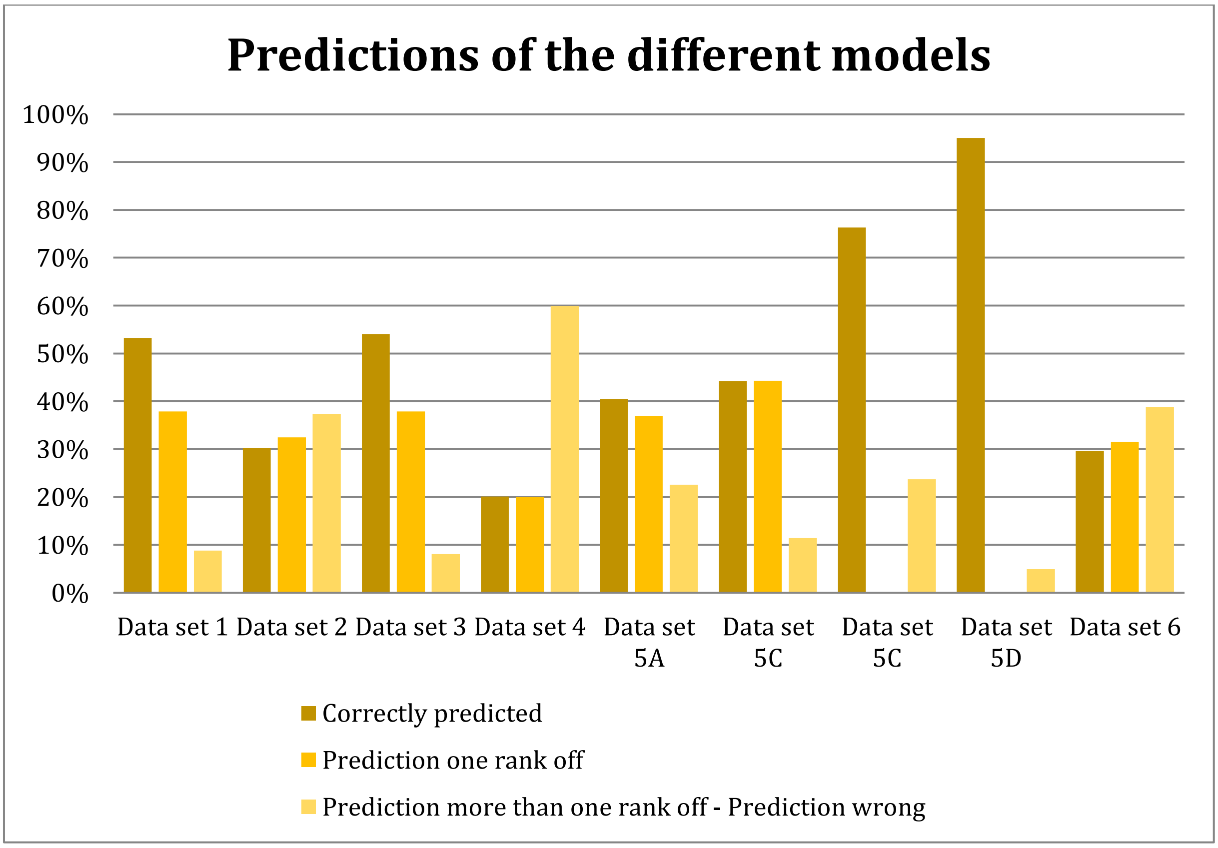

FIGURE 15: DISTRIBUTION PREDICTIONS USING CROSS VALIDATION

4 CONCLUSION

4.1 DOOR OPENINGS

For the absolute number of door openings "duration" seems to be a key indicator for the number of door openings. The longer the operation the higher the number of door openings. Model 1 and model 3 also show that the characteristics day part, start date, end time and operation room have some predictive value for the number of door openings.

4.2 DOOR OPENINGS PER HOUR

For the door openings per hour, the operation room seems to have a good predictive value. Model 2 shows the following distribution of door openings per hour among the operation rooms:

Operation rooms with a very low rank:

- OK 1
- OK 3
- OK 11

Operation rooms with a low rank:

- OK 4
- OK 6
- OK 7
- OK 13

Operation rooms with a medium rank:

- OK 5
- OK 10
- OK 12

Operation rooms with a high rank:

- OK 2
- OK 8
- OK 9

Operation rooms with a very high rank:

- OK 14
- OK 15
- OK 16

The distinction of the different ranks per operating room could be caused by the fact that certain operations usually take place in the same operating room.

By differentiating between the different specialisms and the number of door openings per hour the following distinction can be made:

Specialisms with a very low rank:

- ANS
- OOG
- ORT

Specialism with a low rank:

- NCH
- PCH
- URO
- URK

Specialisms with a medium rank:

- HGE
- HLO

Specialisms with a high rank:

- CCH
- MON
- TRA

Specialisms with a very high rank:

- GYN
- HLK
- HON
- KCH
- KNO
- ONB
- URK
- VER

4.3 DIFFERENT DOOR TYPES

By distinguishing the different door types the following conclusions can be made:

4.3.1 DOOR OPENINGS: SETTING ROOM

Even though the model that predicted the number of door openings of the setting room seems much more elaborate than the model for the combination of all the doors, it is in general quite similar. The number of door openings of the setting room seems mostly dependent on the duration of the operation although start date, end time and room also seem to have some predictive value.

4.3.2 DOOR OPENINGS: SCRUB ROOM

Since the largest percentage of the door openings are door openings of the scrub room, as showed in Figure 4, it is not unusual that these predictions are similar to the predictions for all the doors combined. The number of door openings of the scrub room seems mostly dependent on the duration of the operation although start date, and room also seem to have some predictive value.

4.3.3 DOOR OPENINGS: PATIENT ENTRANCE

The total number of openings of the patient entrance is very low compared to the number of openings of the doors to the setting room and the scrub room, as shown in Figure 4. Model 5C showed that only for room OK-1 it is predicted that the door to the patient entrance will open.

4.3.4 DOOR OPENINGS: PATIENTIENT ACCOMPANIST ENTRANCE

As well as the number of openings of the patient entrance, the number of openings of the patient accompanist entrance is relatively low. Model 5D has shown that only the operation code seems to have predictive value for the number of openings of the patient accompanist entrance.

5 DISCUSSION

To completely stop movement within a theatre complex is impossible, since there is always a possibility of an emergency situation. However, movement can be minimized by monitoring. As an example, previous studies have investigated the reason for door openings and have shown that the single greatest contributor are requests for information; whether to ask questions, to check on cases, or to process paperwork.[5] Measures can easily be taken to prevent these kind of door openings, like the encouragement of means of communication that can be used from inside of the operating room for instance using a phone instead of speaking face to face.

The data used in this study contained operations that took place between June 2011 and February 2012. At the VU medical center monitoring door openings of orthopaedic operations was initiated in March 2012. The average number of door openings for this specialism is displayed in Figure 16. A considerable drop in average number of door openings in orthopaedic operations is shown since monitoring is initiated. An advice for the VU medical center is to continue with monitoring, starting with the specialisms or operation rooms that showed a (very) high rank in door openings.

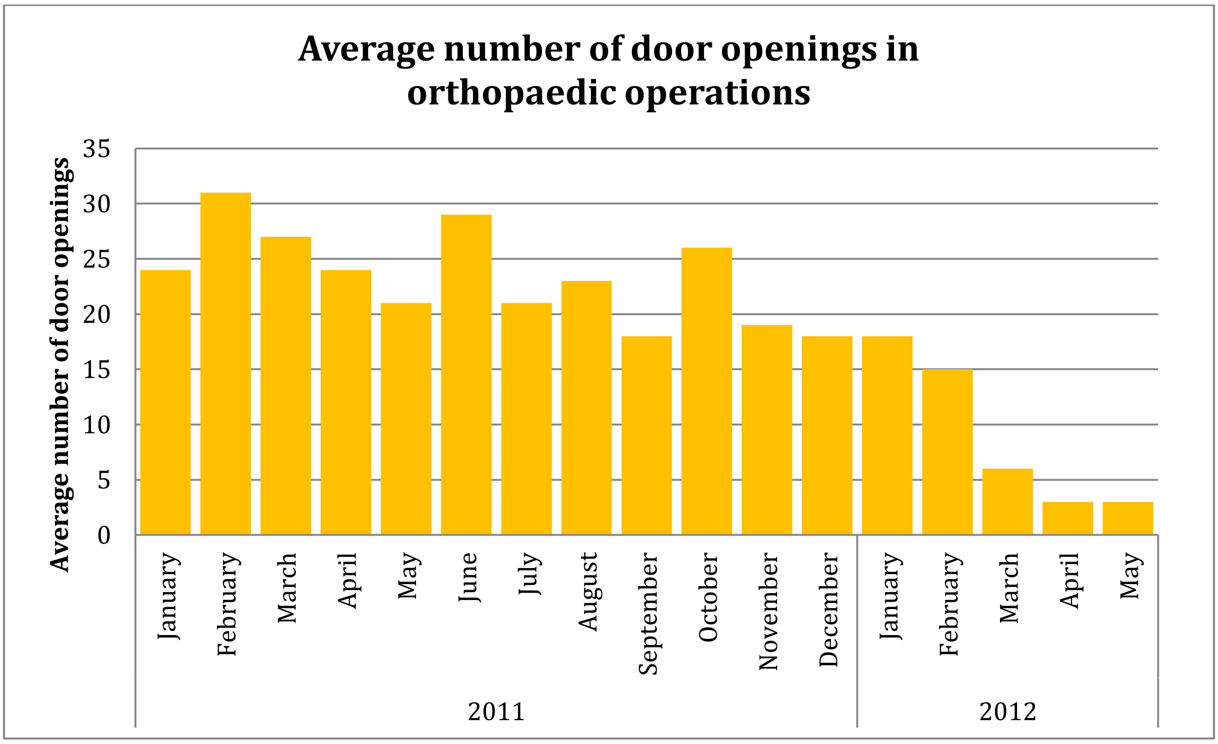

FIGURE 16: AVERAGE NUMBER OF DOOR OPENINGS IN ORTHOPEDIC OPERATIONS

6 FUTURE STUDIES

This study was blinded from patient identifier data, which prevents direct comparison of door openings to SSI rates. Even though it has been previously established that operation theatre traffic is associated with a higher postoperative infection rate, it is recommended to conduct the same research including SSI rates to identify more possible correlations.

7 REFERENCES

[1] Young R S, O'Regan DJ: Cardiac surgical theatre traffic: time for traffic calming measures?: Interactive Cardiovascular and Thoracic surgery.
[2] Dharan S, Pittet D: Environmental controls in operating theaters. Journal of Hospital infection
[3] Young R S, O'Regan DJ: Cardiac surgical theatre traffic: time for traffic calming measures?: Interactive Cardiovascular and Thoracic surgery.
[4] Nelson D: Dictionary of mathematics.
[5] Lynch RJ, Englesbe MJ, Sturm L, et al: Measurement of foot traffic in the operating room: Implications for infection control: American Journal of Medical quality.