Reduction of the airway resistances during lung aeration of the preterm rabbit

Clients
Dr. Janneke Dekker
PhD student Fleur Brouwer

Consultants
Levi Duijst (s1863118) & Alexandros Ioannou (s3630129)

Tutors
Prof. Dr. Hendriek Boshuizen & Dr. Fred van Eeuwijk

Abstract

Lung aeration for infants goes through a transformation starting from being in the womb and relying on the mother's oxygen supply to breathing by itself at birth without any support. For preterm infants this lack of support is especially detrimental to their long-term growth and a solution via the use of controlled ventilation is most suitable. This study aims to look at the use of various settings for ventilation and specifically which combination of them reduces airway resistance in lung aeration. A total of 105 rabbits were used over multiple experiments, where different groups of rabbits were set to different levels of settings. To analyze the data, we took a majority voting approach applying both K-Means clustering in tandem with Bagging taking into account all settings, and a Linear Mixed Model where we applied a random intercept to take into account the mother effect of the rabbits. From the results of these models, we conclude that the settings Target Vt/Kg at level 10 and Rate at levels 24 and 30 help minimize the resistance and time to minimal resistance the most.

Keywords: Lung aeration, Rate, Target tidal volume, ventilation

Introduction

Breathing is an essential part of life. From the moment of birth, animals exhibit breathing as a survival mechanism such that they can grow and mature. Looking at humans closer, breathing starts when the fetus is still in the womb of their mother. Breathing is not aimed at gas exchange, because the fetus receives oxygen via the umbilical cord. Rather, during these breathing movements, the lungs are filled with fluid, and small volumes of liquid move in and out of the airways which enables the lungs to develop. After birth, the cord is cut and the fetus can no longer rely on the mother for oxygen. The liquid present in the lungs of the fetus needs to be cleared to enable aeration of the lungs and subsequently oxygenation of the body. This liquid cleared from the airways moves to the lung tissue and clears soon thereafter via circulation and lymphatics[1]. Most infants adapt to this transition smoothly, but in some cases, assistance is required as in the case of preterm (early) infants with immature lungs and respiratory systems.

At birth, infants must rapidly adapt from the liquid stage in the womb to breathing independently in a gaseous environment, without relying on their mother's oxygen supply[1]. This switch from liquid to gaseous stage causes the lungs to change physically into three stages: 1) lung aeration, 2) liquid accumulation in lung tissue, and 3) liquid cleared from the tissue. During lung aeration, the lungs go in transition from placental gas exchange to pulmonary gas exchange[1]. Therefore the airways must be cleared from liquid. Most infants undergo the physiological changes smoothly, however, preterm infants require more assistance at birth during this stage of life because they are too immature. Their underdeveloped lungs, high chest wall compliance, and weaker inspiratory muscles than their term counterparts, make them struggle to clear the airways of liquid[2].

In contrast, preterm and near-term-born caesarean infants have a combination of lung immaturity and surfactant deficiency which causes them Respiratory Distress or transient tachypnea of the newborn (TTN). As the mothers had a caesarean operation, the infants are not prone to clear liquid via nose and mouth, as they did not undergo postural changes that occur during labor and vaginal delivery, but rather solely via lung tissue absorption. This results in increased liquid repositories which causes the expansion of chest walls and flattened diaphragm[3]. That is why preterm infants need neonatal resuscitation to help them go through this difficult changing process of the lungs[1].

Furthermore, for preterm and near-term caesarean section-born infants, the main consequences are pulmonary edema which in adults tends to cause tachypnea and dyspnea. For preterm infants, lung injuries due to immaturity in the muscles vary and are quite broad which include the aforementioned problems as well as failure to develop functional residual capacity after birth, weak inspiration, and high chest wall compliance[1]. Moreover, to treat the problem at hand, only one solution currently is known. As breathing is very taxing, a tube where ventilation can be controlled is the ideal solution. However, this is a very invasive approach to the infant, and the method can cause damage to the lungs. Additionally, it is necessary to set specific ventilation settings that would stimulate the problem of breathing for them in a much simpler, less taxing manner so that their muscles can mature at their own pace as settings can be changed via the machine[1]. The latter in particular is a case-by-case usage and not a generalized approach.

The main focus of this paper is the initial phase of lung aeration in infants. Liquid in the lungs causes airway resistance to significantly increase (100-fold) and makes it harder for air to enter the lungs and perform gas exchange. To establish effective respiration, this resistance should be minimised as fast as possible. The primary mechanism for airway liquid removal after birth is hydrostatic pressure that is generated by inspiration[1]. Airways resistance decreases during this phase when the lungs are filled with air, however, this reduction follows an exponential function that is difficult to predict[1].

Historically speaking, there is an application of respiratory support in combination with chemical compounds to help preterm rabbits better where it was determined that only respiratory support could in reality make the difference as well as surfactant[2]. In another paper, it was shown that the change in ventilation settings for one parameter (positive end-expiratory pressure) had a great effect on near-term caesarean section-born rabbits[3]. Most preterm infants need neonatal resuscitation at birth to adequately aerate their lungs. However, this assistance should be gentle to avoid further damage to their underdeveloped and vulnerable lungs. In this research, various ventilation settings were tested to determine which strategy works best to reduce airway resistance during the initial phase of preterm caesarean section at birth. The main research question is: which combination of ventilatory settings leads to the quickest reduction in resistance at birth?

Method

Experimental set-up

The dataset consists of multiple experiments, fifteen in total, that have been performed at the synchrotron in Japan between 2010-2018. The set-up for each experiment has been kept the same throughout the years. Rabbits were used for this experiment because rabbits are phylogenetically closer to humans than other rodents, therefore their airway anatomy resembles that of those of humans, which makes them a perfect candidate for lung research[4,5]. Furthermore, rabbits have a short gestation period of 29 – 33 days[6]. The preterm rabbit kittens (*Orychtolagus cuniculus)* were delivered by caesarean section, intubated before they could take their first breath, moved to the x-ray beam, and attached to the ventilator. From the moment the ventilation started, x-ray images at high frequency were gathered at the same time so that aeration of the lung could be visualized. For two minutes, the resistance in the lung due to the presence of the liquid was measured to see how fast the resistance decreases of lung aeration during this time frame in different ventilation settings. In each experiment, only the control rabbits for this study were included. In other words, the chosen rabbits/kittens did not undergo any form of intervention or receive additional treatment that could have influenced the aeration of the lung. Each experiment consisted of 5-8 rabbits, except group 2 which had a total of 14 rabbits and each kitten had a particular indicator for the number of siblings indicating that they shared the same mother. This brings in a total of 105 observations in the dataset.

Settings for lung aeration

With the ventilation machine, seven settings could have been switched between different levels. Each of the fifteen groups had its ventilation settings, which means that fifteen different combinations of parameter settings were tested. The seven settings are explained in Table 1 with all levels available.

Table 1 The seven variable settings with the number of levels and meaning of each setting explained.

Setting	Levels	Meaning
Target tidal volume	4, 5, 8, 10	Amount of air that is inhaled/exhaled during a single breath in mL/kg
Respiratory rate	24, 30, 50, 60, 100	breaths per minute
PEEP (positive end-expiration pressure)	0, 4, 5, 8	Positive pressure with regard to atmospheric pressure in the airway at the end of expiration in cm H_2O[7]
PIP (peak inspiratory pressure) at start	20, 25, 35	maximum pressure at the start of ventilation in cm H_2O
Ti/Te	1.0/1.0, 1.0/1.5	the ratio between inspiratory time (Ti) and expiratory time (Te)
Surfactant present	yes, no	A complex mixture of specific lipids, proteins, and carbohydrates acts to decrease surface tension at the air-liquid interface of the alveoli[8].
Expiratory resistance	low, medium, high	Resistance in the lung that is experienced when exhaling

The setting combinations that have been chosen were based on clinical relevance, which is why not all combinations were used (too many experiments should have been performed) and why certain settings appear more often than others.

Measurements

The data contained two main output variables: time to minimum resistance in the lung and the minimum resistance in the lung. Time to minimal resistance is the time point in seconds after ventilation at which the resistance is at a minimum and does not change anymore, in other words, the resistance in the lung remains constant after that time point. The minimum resistance in the lung is the value of resistance in cm H_2O/mL/kg at the time point at which the resistance is at a minimum. Lung aeration was thought to consist of two jointly distributed responses, where both responses were treated by themselves. Not only is it needed to reach the minimum resistance in a short time frame, but the resistance itself has also to decrease substantially in the lung to make sure that the rabbit kitten can breathe optimally. To obtain both the time to minimum resistance and the minimum resistance, an algorithm was used to find these values for each kitten. First, the data was smoothed, with time in seconds on the x-axis and resistance in H2O/mL/kg on the y-axis. Then the first derivative (speed) and the second derivative (acceleration) were calculated. The point where both the first and second derivatives intersect is the point where the difference is the smallest between the derivatives. From the moment the descending resistance started till the moment the derivatives intersected, the smallest distance between the first and second derivatives was determined, which is the point of minimal resistance with corresponding time to minimal resistance (F. Brouwer, personal communication, November 14, 2023), see also Figure 1 in the Appendix. Besides these two outcome variables, the birthweight in grams of each preterm kitten was measured and the gestational age (GA) was noted, which was either 28 or 30 days. Gestational age estimates how far along the kittens are during the pregnancy from the last menstrual cycle of the mother till the caesarean section.

Statistical analysis

For the statistical analysis, two methods were performed namely K-means clustering combined with Bagging and linear mixed model. The main goal of this report was to associate the level settings to the outcome variables minimal resistance and time to minimal resistance. Because there is no direct connection between the explanatory and outcome variables because there were two outcome variables that were needed for the analysis. Both minimal resistance and time to minimal resistance are important factors that needed to be included in the analysis to make inferences about the relationship between the variable settings. A clustering program can be used to summarize both outcome variables and where each kitten is placed into a cluster with other ones similar to it concerning minimal resistance and time to minimal resistance.

K-means was picked because the data looked linearly separable with clear clusters of variable size, easy to interpret as each of the observations is placed in one and only one cluster and each cluster can be analysed separately to gain insights[9]. K-means clustering is a tool for finding cluster structure in a data set that is characterized by the greatest similarity within the same cluster and the greatest dissimilarity between different clusters[10]. K-means is an unsupervised learning approach, so the algorithm learns without unlabelled data. Once the optimal number of clusters is established, the K-means algorithm assigns each data point to a specific cluster, minimizing the sum of square distances within each cluster. From the K-means model, the group number for each kitten can be obtained and used as an outcome variable. This new variable is a group number that showed in which cluster each kitten belonged. From data exploration it was found that three clusters looked like the optimal value with a cluster of interest, see Figure 2 in the Appendix, so K-means clustering is suitable for this dataset.

The next step is investigating which variable setting level led to the "cluster of interest". This cluster of interest contains kittens that had both a low time to minimal resistance and low minimal resistance. After the kittens were placed in their respective cluster, the next step was to determine which of the seven settings determines that an observation is placed in a cluster. In other words, the problem was transformed into a classification problem where the main interest is which predictors lead to the cluster of interest. This was done with a tree-structured classification approach that is not based on assumptions of normality. Bagging was picked because it is a technique that reduces the variance and overfitting associated with prediction and to obtain a measure of feature importance; thereby improving the prediction process by bootstrap aggregating[11]. From the prediction model, the importance of each variable can be retrieved to see which variable had the most influence in classifying the kittens in the cluster of interest. All data was used for both K-means clustering and bagging. One drawback is that the K-means and bagging approach do not account for dependence between siblings, nor does it account for the correlation between the kittens.

To include these assumptions in the approach, a Linear Mixed Model was fitted with a random intercept for kitten numbers to deal with the same mothers of rabbits. In this way, we can aptly handle the within-kitten sibling groups and between-kitten sibling group variability. As we are dealing with repeated measurements for each kitten, a diagonal correlation structure was chosen to produce accurate inferences for the relationship within each group.

The reason why two approaches were used is due to the idea of majority voting, whereby various models were applied to a classification problem and the result chosen is based on the agreement of the majority of models. For this study, suppose there is a consensus between the two approaches and the results in both cases are in agreement then this is beneficial because the weaknesses of one model can be offset by the strengths of the other model, making the overall conclusion more robust. In case of dissonance then an in-depth analysis of why each model performed differently would be prudent and perhaps one of the models would not be beneficial for this particular problem. All statistical analysis was done in R (Version 4.2.2).

K-means clustering

The two outcome variables time to minimum resistance and minimum resistance are both important in explaining which ventilation settings are important for lung aeration because both a short time should be reached and the resistance in the lung should be as low as possible. K-means clustering works best with two outcome variables due to good inference as in this case, no dimension reduction technique was needed and ensures that both variables remain independent in the analysis. The idea of K-means clustering, an unsupervised learning technique, is to group the data points in clusters as well as possible by minimizing the within-cluster variance (how much the data points deviate from each other inside a cluster) of all clusters as well as possible. The method starts with random clusters and continues until the minimum within-cluster variance is the lowest. Observations that are most similar to each other are clustered together into one group. The parameter for K-means is the number of clusters and this needs to be given. The K-means algorithm runs for a range of clusters. Multiple tests are available to check how many clusters are needed. Two such tests are the average silhouette and the within-cluster sum of squares[12]. Both of these tests try to find the optimal number of clusters given the data by having certain criteria. The average silhouette method looks at the cohesion and separation of data points in and between clusters, respectively. The value ranges between -1 and 1. The meaning of 1 means that a data point is far from neighbouring clusters, so there is a clear separation of the clusters. The within-cluster sum of squares measures the variability of the data points in each cluster. The smaller this variability, the more compact the clusters. The trade-off with this method is between the lowest cluster sum of squares against a small number of clusters for interpretability. To find the optimal number of clusters that have a low within the sum of squares and can still be easily interpreted, the optimal number of clusters is found at the "elbow" of the graph. If both of these methods have a consensus on the number of clusters, then it is most likely that many clusters are optimal for the dataset. Another important aspect of k-means clustering is that the data is scaled before doing the clustering technique because the two outcome variables were measured in different units and each unit has its scale. For example, a difference in 1 second is less than a difference in 1 cm H2O/mL/kg. When scaling data, the range of the variables is changed to be equal to each other and no variable with a higher range has more influence on the clustering. The K-means function is available in base R. For a visual representation, the package factoextra was used.

Bagging (Bootstrap aggregating)

The method of bagging or bootstrap aggregating is an ensemble learning method. It can work as a classifier where it tries to predict in which cluster each observation belongs based on the explanatory variables. The most important feature of bootstrap aggregating is the decision tree. The bagging part is to prevent overfitting of the model by making multiple classifiers and combining the predictions based on the majority of votes. What bagging does is make new datasets based on the original dataset through random sampling with replacement, so you take a random kitten from the dataset and add it to the new dataset. This means that a kitten can be present once, once, or multiple times in a new dataset. Each new dataset is modeled with the same predication method and their combined predictions are aggregated together, by averaging for simple voting for classification, to obtain the overall prediction[11]. Another reason why sampling with replacement is useful is when the class distribution is unbalanced. Bagging solves this problem with random sampling, where the class distribution can be evenly balanced to prevent bias of the majority class and improve the performance of the minority classes.
The model that was used for classification was a decision tree. A decision tree is a flowchart structure for class labeling. The tree consists of three components: a node, a branch, and a leaf. The node can be seen as a test for which it can be answered with a yes or a no. Each node/test is based on one of the exploratory variables that is placed in the model; e.g. Is Target VT/kg = 10? The branches are the decision or answers that have been made based on the test and leads to another node (test). The leaf is the final point of the tree and represents a class label. To fit the bagging model the package randomForest was used.

Linear Mixed Model

This model ensures correlation, clustering, and unbalanced data are considered and ensures greater accuracy. For the use of this model, one outcome variable had to be selected, combinations of the two were also considered but ultimately resorted to uninterpretable results. Recognizing that constraining Minimal Resistance to be below 600 creates an ideal scenario with data points exhibiting the required level of resistance, the focus shifts toward utilizing the time to minimal resistance as the primary outcome variable.
Linear Mixed Models (LMMs) are a type of supervised learning technique that extends linear models. Unlike traditional linear models that assume ideal situations with trends derived by minimizing errors, LMMs are designed to handle more complex and imperfect scenarios in data analysis, where correlations can impact the results. They incorporate both fixed effects, which

represent factors affecting the entire population, and random effects, which capture individual variability. Additionally, they include correlation matrices to account for inherent correlations within individuals or observations. LMMs can be decomposed thus to three components, the fixed and random effects and the correlation matrix.

All the settings in the experiment analysed were considered fixed effects and the sibling factor, kittens with the same number of siblings coming from the same mother were considered the random effect and in particular the random intercept. We suspect that this factor enables observation dependency and consequently correlation between rabbits, for example, a kitten with 5 siblings would be similar to its siblings biologically speaking.

The errors were also modelled via a correlation matrix to account for the unbalanced data and sibling relationships, in particular, a diagonal matrix was considered most suitable as it deals with members of the same family. Assumptions for this model include the random effects and error terms following a normal distribution and also the error terms independent of the covariates. The alpha level was set at 0.05 and the R package nlme was used.

Results

<u>K-means clustering</u>

Based on the silhouette plot and within the sum of squares (see also Appendix Figure 3), the number optimal number of clusters was three. So K-means clustering has been done with three clusters and the results can be seen in Figure 4.

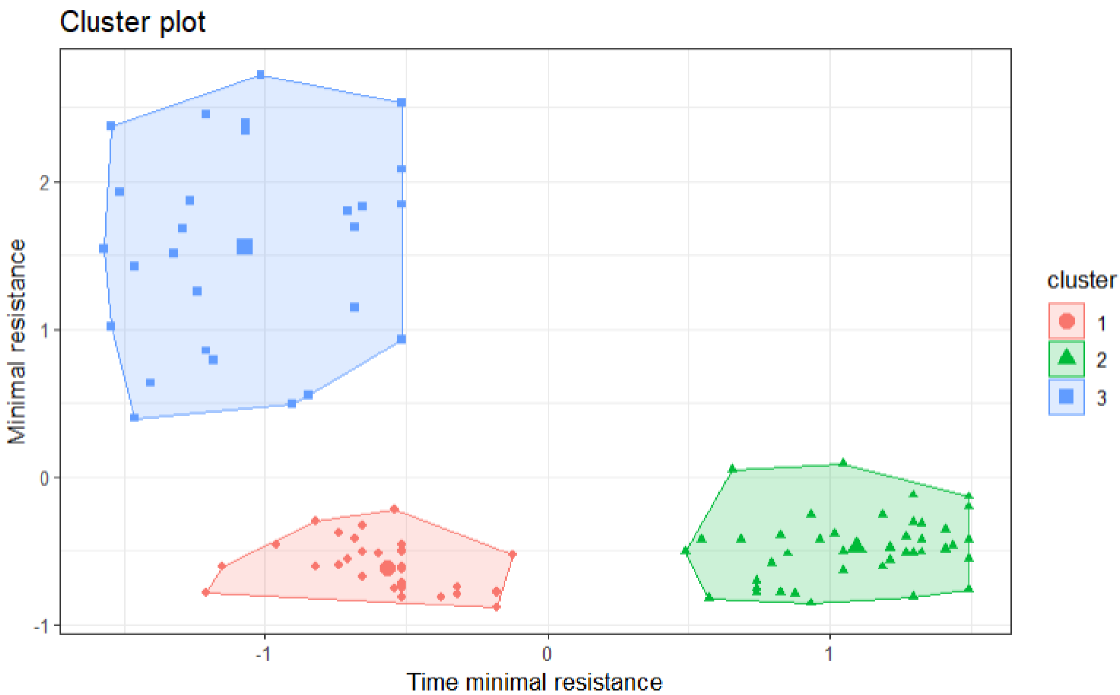

Figure 4 Based on the two variable outputs time minimal resistance and minimal resistance three clusters were based on the silhouette plot and within the sum of squares plot. All three clusters do not overlap each other. The red cluster 1 is the one of

interest because it contains the kittens with the device settings that lead to a fast decrease in resistance and the resistance in the lung is relatively low compared to cluster 3 (blue cluster).

The three clusters did not overlap with each other. In Table 2 an overview is given of the different settings that were presented in Cluster 1, but not in Cluster 2 and 3.

Table 2 Overview of the different level settings that were present in Cluster 1 (red), but not in Cluster 2 and Cluster 3.

Setting variable	Present in cluster 1 but not cluster 2	Present in cluster 1 but not cluster 3
Target tidal volume	10	10, 8
PEEP	/	4
PIP at start	/	25, 20
Respiratory rate	24, 30	30, 60
Expiratory resistance	High, Medium	High, Medium
Ti/Te	1.0/1.5	1.0/1.0
Surfactant present	/	/

From Table 2 it can be noted that certain setting levels were only present in cluster 1 namely: target tidal volume of 10 VT/kg, respiratory rate of 30, and expiratory resistance with high and medium levels.

Bagging

With the bagging method, all data was used that was available. Bagging cannot account for correlation between variables and if correlation is present in the model, then most decision trees will look similar to each other. It was stated before that rate and Ti/Te are correlated **(**see Figure 5 in the Appendix)**.** Because these two variables were correlated, they explain the same information; when Rate increases Ti/Te decreases and vice versa. The consequence of this was that Ti/Te importance was set very low and lost most of its importance. Furthermore, GA was also added to the bagging method.

The bagging model returned two important plots: the variable importances, which showed which variable had the most influence on the class labeling of the kittens. From Figure 6 it can be observed that Rate was the best-performing variable in the decision-making. This means that when the rate is left out of the model, the accuracy will decrease by 100%, which is equal to randomly distributing the class labels to the kittens. The next important variable was Target tidal volume with a mean decrease in accuracy of 83%. PEEP is the third important variable with only 9% importance. The other five variables had a very low mean decrease in accuracy, which means that these variables were less important in the decision-making process for classification.

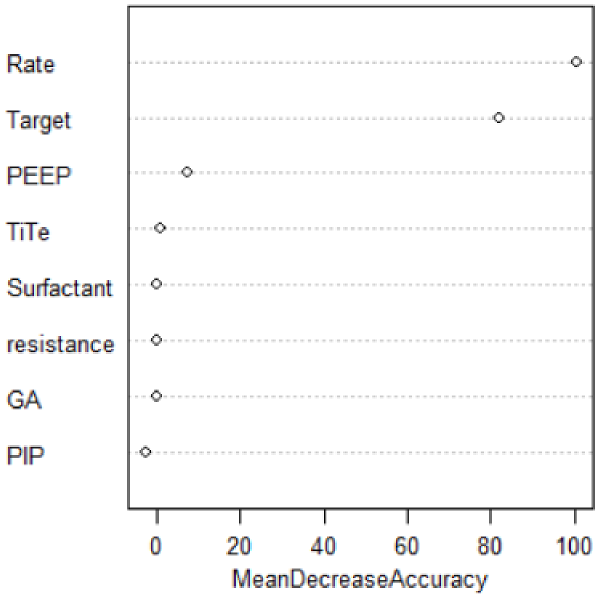

Figure 6 Variable importance plot of the five variables in the bagging model. From the plot, it can be seen that target tidal volume is the most important variable in the model, because when removing this variable from the model. The accuracy decreases by 100%, which is equal to random labelling of the classes to the kittens. The other variables have a lower mean decrease accuracy and are therefore less important for the classification.

Next, the partial dependence plots show which of the levels of the variables was more likely the cause of placing a rabbit kitten in the cluster of interest/cluster 1 (see Figure 7). This could also be shown for the other 2 clusters, but these were not the main interest of this study. Based on these plots, the most likely levels in the cluster of interest can be subtracted. The partial dependence plot for Rate shows the log odds (the higher the log odds the higher the probability of being in cluster 1) for rates 24 and 30. This means that when the variable setting of Rate was set at either two of these values, that would have caused the classification of a rabbit kitten to be placed in cluster 1 (low minimal resistance and low time to minimal resistance). When looking at the partial dependence plot on Target, it clearly shows that a target of 10 VT/kg was most likely to result in cluster 1. The same can be applied to the other variables. For PEEP it is level 4 that is most likely to result in a kitten ending up in cluster 1 and a Ti/Te of 1.0/1.5 too. When a value is negative in the partial dependence plot, that indicates that it is less likely for a kitten with that level to be placed in cluster 1, or in other words; the kitten will have high minimal resistance or a high time to minimal resistance.

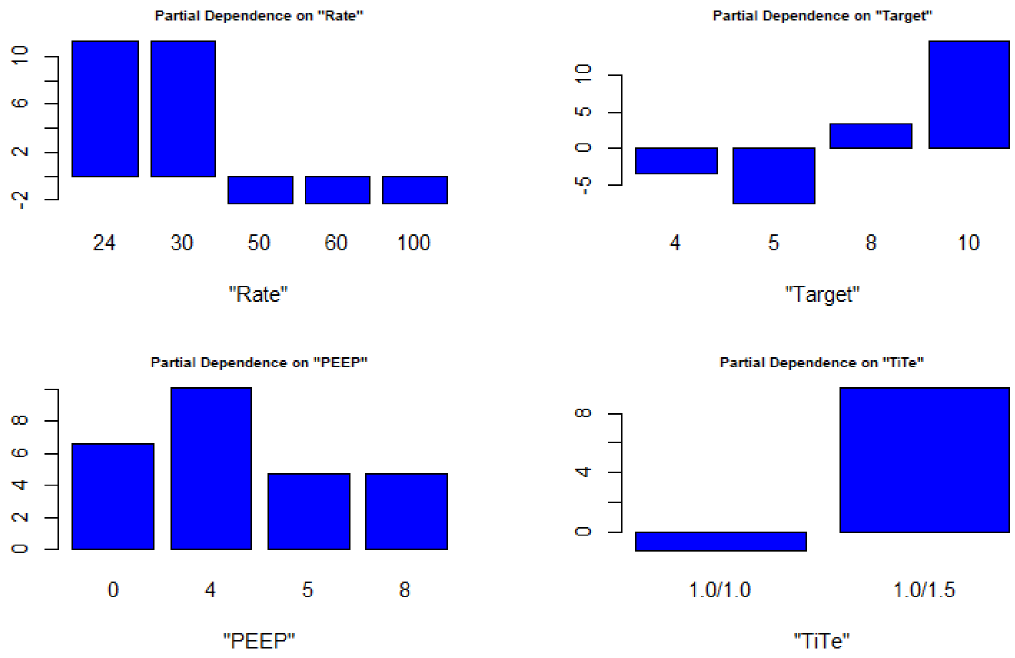

Figure 7 Partial dependence plots on the four variables: Rate, target, PEEP, and Ti/Te. The x-axis represents the levels for each variable and the y-axis is the log odds of being present in the cluster of interest (cluster 1 in this case). So how higher this value, the more likely it is that this level of the variable leads to being placed in cluster 1. For the target tidal volume, it is level 10 VT/kg that is most likely to be in cluster 1. A PEEP of 4 and a PIP of 35 leads to be placed in cluster 1. For surfactant present and expiratory resistance, all of these levels had a similar probability of being placed in cluster 1.

The other four variables' partial dependence plots can be found in the Appendix (Figure 8).

<u>Linear Mixed Models</u>

By setting a boundary on Minimal Resistance smaller than 600 and removing Rate and Ti/Te as predictors in our model due to collinearity a linear mixed model was fitted. Setting the correlation matrix to be independent to take into account the correlation between sibling kittens and setting the kitten number as the random intercept term the estimation of the level variables can be observed from Table 3.

Table 3 Regression table including the significance level for each predictor. The intercept, PEEP5, PIP35, and TargetVtkg10 are significant with a p-value smaller than 0.05. Rounded to 2 d.p. All variables are categorical with Target Vt kg 10 and

Surfactant Presence also being dummy variables taking only 1 and 0 as values. Target VT Kg is a dummy due to bounding Minimal Resistance to less than 600 and only Target Vt kg at levels 8 and 10 being present henceforth in the filtered data.

Settings	**Coefficients**	**Standard Error**	**P-value**
Intercept	81.61	19.89	<0.01
PEEP 4	19.39	13.00	0.141
PEEP 5	21.27	9.07	0.022
PEEP 8	19.31	10.54	0.072
PIP at start 25	-0.44	12.78	0.972
PIP at start 35	-17.30	7.72	0.029
Target Vt kg 10	-42.70	7.72	<0.01
Expiratory resistance Low	4.13	10.26	0.689
Expiratory resistance Medium	0.53	14.71	0.972
Surfactant Present Yes	4.16	9.07	0.648

From Table 3, PEEP with level 5, PIP at the start with level 35, and Target Vt/kg with level 10 were significant with p-values lower than 0.05. They had coefficients of 21.27, -17.30, and -42.70, respectively. All other variables were not significant as can be seen from the p-value larger than 0.05. The standard error is quite large for these predictors as well, this is due to the extreme variable observations as can be seen from the two clusters having very large spread of values.

The intercept included PEEP with level 0, PIP at the start with level 20, Target Vt/kg with level 8, no surfactant presence, and expiratory resistance set to high. The coefficient for the intercept is 81.61 seconds to reach the time to minimal resistance and is the reference level. The coefficient for PEEP with level 5 is 21.27 meaning that this setting increases the time to minimal resistance.

On the other hand, the coefficient for PIP at the start with level 35 has a coefficient of -17.30, this means that by applying this setting the time to minimal resistance goes down by -17.30 seconds. This is also the case for Target Vt/kg with level 10 with a coefficient of -42.70 whereby this setting reduces the time to minimal resistance compared to the reference level.

From these settings, the Target Vt/kg with level 10 provides the strongest decrease in the time to minimal resistance. The inclusion of a Target Vt/kg of 10 on the intercept reduces the time to a minimal resistance of 38.91 seconds.

The random intercept was negligible as each kitten's sibling number coefficient was very small. This indicates that within each sibling group, no significant differences were observed implying that each group has the same intercept.

Mathematical formulation of the LMM model, further example calculations, analysis of the assumptions of the LMM, and random intercept are provided in the appendix.

Discussion

The research question of this paper was which combination of ventilatory settings leads to the quickest reduction in resistance at birth? The clustering approach with K-means together with bagging leads to two important variables: Rate and Target Vt/Kg. Of these variables, the level settings of a Rate of 24 and 30 and Target Vt/Kg of 10 had the highest probability of classifying the kittens in cluster 1, which is the cluster that minimises both minimal resistance and time to minimal resistance. The linear mixed model showed that the coefficients for Target Vt/Kg of 10 and PIP at the start of 35 were the level settings that reduce the time to minimal resistance significantly. Both approaches reached the same conclusion about Target Vt/Kg of 10, but PIP at the start had a very low importance level from the bagging model, contradicting the linear mixed model. Rate and Ti/Te were not included in the LMM due to collinearity.

Because of this, other variables could take up the importance of explaining the outcome variable when Rate and Ti/Te were not present in the LMM. However, the Rate was important in the bagging model and should therefore be included in the final decision. In the end, Rates with 24 and 30 bpm together with a Target Vt/Kg of 10 lead to the quickest reduction in minimal resistance in the lung in a short time frame. Both of these combinations were present in group numbers 7, 8, 9, and 10, see also Table 4, which also indicates that both variables can be combined, it is not an impossible combination.

Both approaches have their limitations. The K-means clustering with bagging did not account for dependencies and correlations between kittens, which can influence the clustering algorithm as kittens with the same mother could result in similar outcome values. Moreover, due to the collinearity between Rate and Ti/Te, the bagging model picked one variable for importance and gave the other variable very low importance. Also, a mean decrease in accuracy of more than 100% indicates that the model was overfitting the data. However, the LMM can deal with correlation by adding it as a random effect. In this model, some data had to be removed as the Minimal Resistance was bound and some parameters were correlated and had to be removed as well, to ensure good inference on the estimators. As a result, the LMM had less data to explain the connection between the outcome and the remaining variables.

Furthermore, the dataset itself was small (only 105 observations) and very unbalanced (not all combinations were present). Because of this modeling can be quite a challenge and doing good inference is difficult when not all combinations are present. However, due to the nature of the experiment, working with animals, and the number of possible combinations, it is not possible to execute all possible combinations. Given the current data, reasonable inference was possible by using two approaches in an explorative way to find common ground between the two. However, the experiments themselves were part of different research questions that were conducted over several years. One can wonder how consistent the experiments were and if they are comparable over 8 years because, in this big period, the experiments and thus data can be different due to changes that may happen over time. Researchers could be inexperienced in the first few years but over time they become better at using the equipment and are more capable of performing the procedure which results in better outcomes.

Comparing the found variable settings from the two approaches and looking at conventional settings from the literature, some differences are noticeable, see Table 5. Firstly, the Target value of 10 is higher than the conventional value setting that is used. However, this level has no association, only higher levels of Target have an association with lung damage, only for higher levels of target tidal volume[13]. The rate variable in literacy has no clear consensus and different levels are used across studies, so this study's findings indicate that a lower rate is good enough than a higher rate. Furthermore, low levels of the rate-setting do not change the CO2 level in stable preterm infants[14], which suggests that it is better to apply lower rates rather than higher ones**.** Of Ti/Te an article stated that a ratio of 1:2 is unnecessary and that a ratio of at least 1:1 is appropriate. A longer Ti improved dynamic lung compliance possibly due to reducing airway resistance[15]. A ratio of 1.0/1.5 is in the middle of the two extreme values. The found PEEP values for both Bagging and LMM were in line with each other and also with the literature. PEEP not being important in the lung aeration phase, does not imply PEEP is not important at all. From the literature, a PEEP value of 4 seems better than lower values like 0 or higher values like 7, because either low or high values can cause lung injuries due to different mechanisms[16]. So PEEP must also be carefully selected for the overall lung aeration process to minimize collateral damage.

Table 5 Final level selection of only the important variables for both bagging and LMM. The third column contains the variable level settings for lung aeration that are found in the literature.

Setting	Bagging	LMM	Literature
Target Vt/kg	10	10	4-6[17]
Rate	24, 30	N/A	10-60[14,17]
Ti/Te	1.0/1.5	N/A	1.0/1.0[15]
PEEP	0,4	0,4,8	4[16]
PIP at start	20	35	15-23[17]

This study only focuses on the first phase of infants that move from the liquid to the gaseous stage during birth. However, there are two other phases after lung aeration. So what is the effect of the found variable settings on the second and third phases; does that cause new problems? The study did not focus on the health characteristics of the preterm rabbit kittens for example if additional health problems arose or if some kittens even died. Health indicators are also important for such studies to understand what the less visible signs are of the ventilation settings than only visible ones. And how well do the kittens do after ventilation: normal life span, no breathing problems, or other behavior changes that are not normally seen in rabbits?

Conclusion

This study focuses on lung aeration of preterm rabbit kittens that were born via a caesarean section. Two approaches were used namely K-means clustering combined with bagging and LMM. Both methods had a consensus on the importance of Target Vt/kg of 10. This variable setting should be included to minimize the minimal resistance in the lung and time to minimal resistance during the lung aeration phase. Furthermore, the variance important plot indicated that Rate is also an important setting variable with rates of 24 and 30 bpm as the best settings that lead to the cluster of interest. A combination of both Target Vt/kg of 10 and a rate of either 24 or 30 bpm in the ventilation setting lead to the fastest and lowest resistance in the lung for preterm rabbits after birth.

Bibliography

1. Hooper SB, Kitchen MJ, Polglase GR, Roehr CC, Te Pas AB. The physiology of neonatal resuscitation. *Curr Opin Pediatr*. 2018;30(2):187-191. doi:10.1097/MOP.0000000000000590

2. Crawshaw JR, Hooper SB, Te Pas AB, et al. Effect of betamethasone, surfactant, and positive end-expiratory pressures on lung aeration at birth in preterm rabbits. *J Appl Physiol*. 2016;121(3):750-759. doi:10.1152/japplphysiol.01043.2015

3. McGillick E V., te Pas AB, Croughan MK, et al. Increased end-expiratory pressures improve lung function in near-term newborn rabbits with elevated airway liquid volume at birth. *J Appl Physiol*. 2021;131(3):997-1008. doi:10.1152/japplphysiol.00918.2020

4. Kamaruzaman NA, Kardia E, Kamaldin NA, Latahir AZ, Yahaya BH. The rabbit is a model for studying lung disease and stem cell therapy. *Biomed Res Int*. 2013;2013. doi:10.1155/2013/691830

5. Bosze Zs. , Houdebine L.M. Application of rabbits in biomedical research: a review. *World Rabbit Sci*. 2010;14(1):1-14. doi:10.4995/wrs.2006.712

6. Rosahn PD, Greene HSN, Hu C. Observations on the gestation period of the rabbit. *Order A J Theory Ordered Sets Its Appl*. 1935;72(1):1-2.

7. Ashbaugh DG, Petty TL. Positive end-expiratory pressure. Physiology, indications, and contraindications. *J Thorac Cardiovasc Surg*. 1973;65(1):165-170. doi:10.1016/s0022-5223(19)40839-8

8. Chakraborty M, Kotecha S. Pulmonary surfactant in newborn infants and children. *Breathe*. 2013;9(6):476-488. doi:10.1183/20734735.006513

9. James G, Witten D, Hastie T, Tibshirani R, Taylor J. An Introduction to Statistical Learning. Published online 2021:612. doi:10.1007/978-3-031-38747-0_10

10. Sinaga KP, Yang MS. Unsupervised K-means clustering algorithm. *IEEE Access*. 2020;8:80716-80727. doi:10.1109/ACCESS.2020.2988796

11. Sutton CD. Classification and Regression Trees, Bagging, and Boosting. *Handb Stat*. 2005;24(04):303-329. doi:10.1016/S0169-7161(04)24011-1

12. Kodinariya TM, Makwana PR. Review on determining of cluster in K-means. *Int J Adv Res Comput Sci Manag Stud*. 2013;1(6):90-95. https://www.researchgate.net/publication/313554124

13. Bland RD, Kim MH, Light MJ, Woodson JL. High frequency mechanical ventilation in severe hyaline membrane disease an alternative treatment? *Crit Care Med*. 1980;8(5):275-280. doi:10.1097/00003246-198005000-00001

14. Hochwald O, Borenstein-Levin L, Dinur G, et al. The effect of changing respiratory rate settings on CO2 levels during nasal intermittent positive pressure ventilation (NIPPV) in premature infants. *J Perinatol*. 2023;43(3):305-310. doi:10.1038/s41372-023-01614-7

15. Pryor EJ, Kitchen MJ, Croughan MK, et al. Improving lung aeration in ventilated newborn preterm rabbits with a partially aerated lung. *J Appl Physiol*. 2020;129(4):891-900. doi:10.1152/japplphysiol.00426.2020

16. Naik AS, Kallapur SG, Bachurski CJ, et al. Effects of ventilation with different positive end-expiratory pressures on cytokine expression in the preterm lamb lung. *Am J Respir Crit Care Med*. 2001;164(3):494-498. doi:10.1164/ajrccm.164.3.2010127

17. Bhat R, Kelleher J, Ambalavanan N, Chatburn RL, Mireles-Cabodevila E, Carlo WA. Feasibility of mid-frequency ventilation among infants with respiratory distress syndrome. *Respir Care*. 2017;62(4):481-488. doi:10.4187/respcare.05157

R Packages used

corrplot: Taiyun Wei and Viliam Simko (2021). R package 'corrplot': Visualization of a Correlation Matrix (Version 0.92). Available from https://github.com/taiyun/corrplot

factoextra: Kassambara A, Mundt F (2020). _factoextra: Extract and Visualize the Results of Multivariate Data Analyses_. R package version 1.0.7, <https://CRAN.R-project.org/package=factoextra>.

ggplot2: H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

nlme: Pinheiro J, Bates D, R Core Team (2022). _nlme: Linear and Nonlinear Mixed Effects Models_. R package version 3.1-160, <https://CRAN.R-project.org/package=nlme>.

randomForest: A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18--22.

readxl: Wickham H, Bryan J (2023). _readxl: Read Excel Files_. R package version 1.4.2, <https://CRAN.R-project.org/package=readxl>.

Appendix

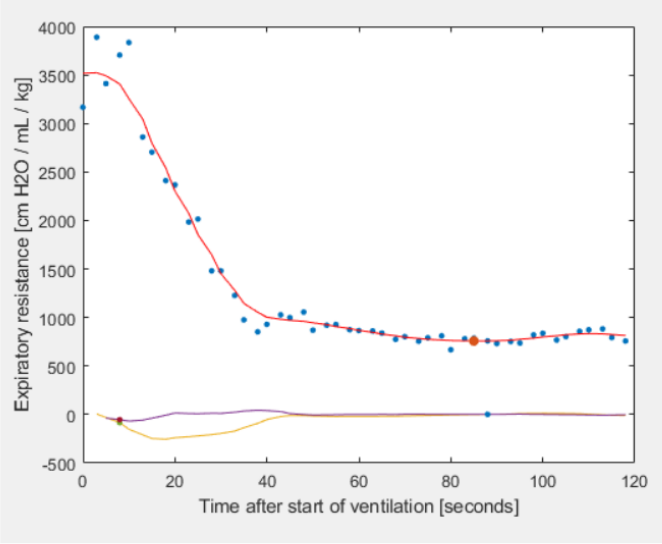

Figure 1 The blue dots represent the resistance in the lung per time in seconds after ventilation. The red line is the smooth line of the datapoints. The first (yellow) and second derivative (purple) are plotted below. The intersection of the 1st and 2nd derivative (blue dot), a moment of descent red dot

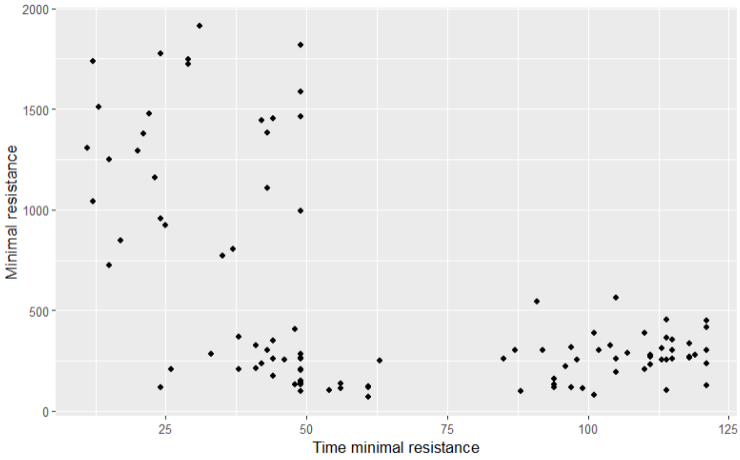

Figure 2 Scatterplot of the two outcome variables Minimal Resistance and Time to Minimal resistance for data exploration. It can be observed from the plot that it was most likely that three clusters were present in this dataset.: left-top, bottom-middle and right-down.

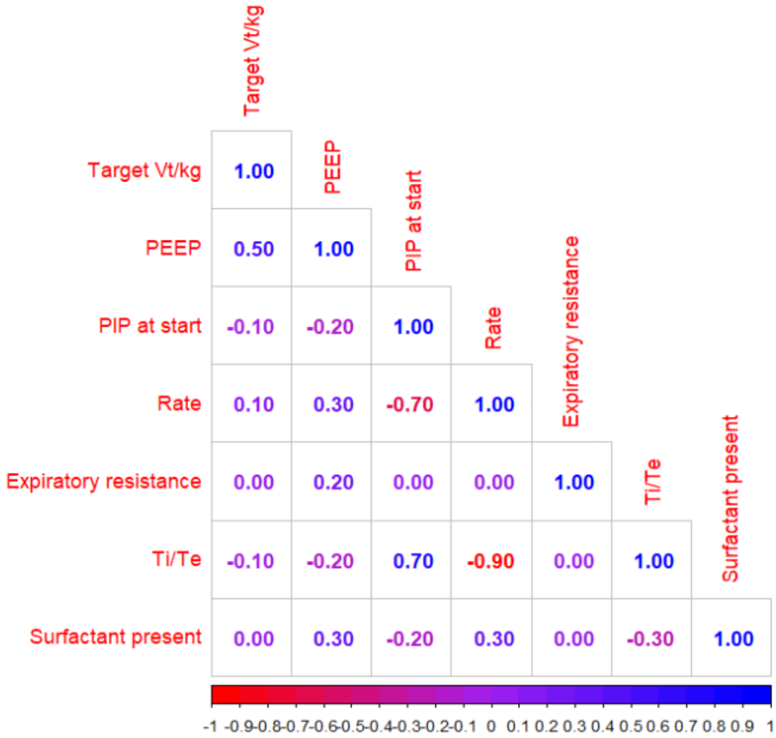

Figure 5 Correlation plot of the seven setting variables. The variable levels were converted to numerical values to make the correlation possible. From the plot, it can be observed that Rate and Ti/Te were negatively correlated. This means that most levels were paired together for most experiments; e.g. not many combinations between the levels of Rate and Ti/Te were present in the experiments. PIP had something similar with Rate and Ti/Te but with a lower correlation. The other variables had a low(er) correlation with the others, which indicated that the level combinations were mixed over the experiments.

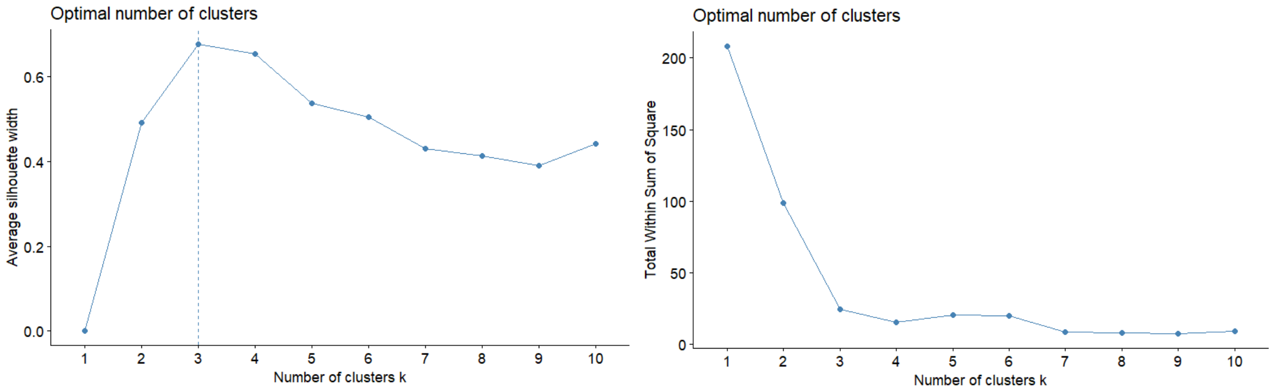

Figure 3 On the left is the Silhouette plot to find the optimal number of clusters based on time minimal resistance and minimal resistance. The optimal number of clusters based on the average silhouette width is 3. On the right plot, the within Sum of Square is plotted against the number of clusters. The "elbow" can be found at number 3. At this point, the within sum of square stayed constant and three clusters is a small number

Table 4 Overview of settings for each group.

Group	GA	Target tidal volume	PEEP	PIP	Rate	Ti/Te	Resistance	Surfactant present
1	28	4	0	35	24	1.0/1.5	Low	No
2	28	5	0	35	24	1.0/1.5	Low	No
3	28	5	5	35	24	1.0/1.5	Low	No
4	28	8	4	35	30	1.0/1.0	Low	No
5	28	8	4	35	50	1.0/1.0	Low	No
6	28	8	4	35	100	1.0/1.0	Low	No
7	28	10	4	35	24	1.0/1.5	Low	No
8	28	10	4	35	24	1.0/1.5	High	No
9	28	10	5	35	24	1.0/1.5	Low	No
10	28	10	5	35	24	1.0/1.5	Medium	No
11	30	8	4	20	60	1.0/1.0	Low	No
12	30	8	0	25	60	1.0/1.0	Low	No
13	30	8	5	25	60	1.0/1.0	Low	No
14	30	8	8	25	60	1.0/1.0	Low	No
15	30	8	5	25	60	1.0/1.0	Low	Yes

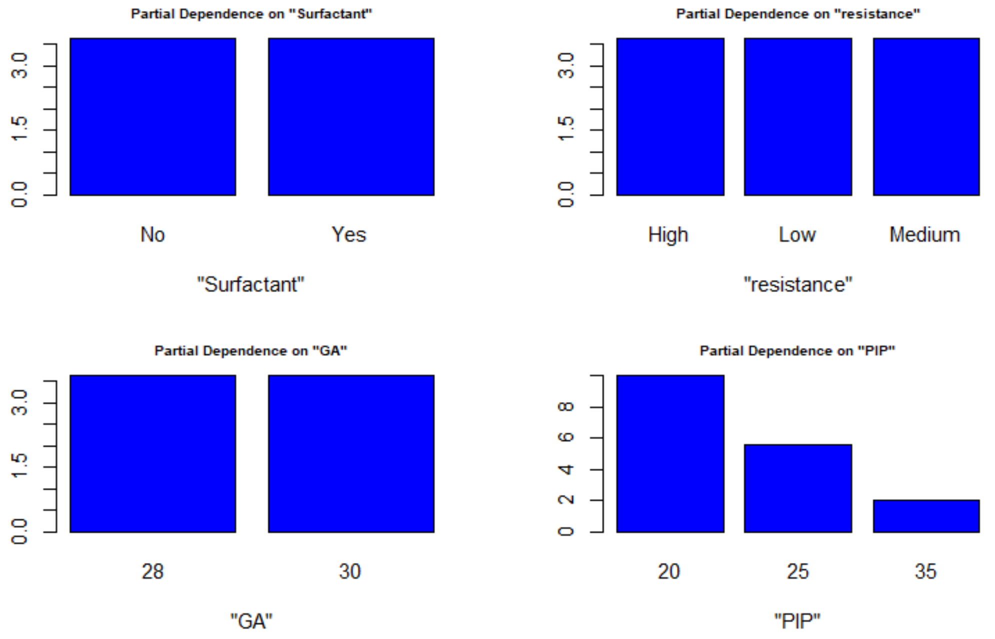

Figure 8 Partial dependence plots of the four variables with the lowest importance based on the variance important plot: Surfactant, resistance, GA, and PIP. The x-axis represents the levels for each variable and the y-axis is the log odds of being present in the cluster of interest. So how higher this value, the more likely it is that this level of the variable leads to being placed in cluster 1. All levels for each variable show a high probability of placing the kittens in cluster 1/cluster of interest.

Linear Mixed Model: Mathematical Formulation of the Regression Table

$$\text{Time To Minimal Resistance}_{i,j,l,k} = \text{Intercept} + \text{PEEP}_i + \text{PIP at start}_j + \text{Target Vt kg 10} + \text{Expiratory Resistance}_l + \text{Surfactant Presence} + \text{KittenNr}_k$$

Where i = {4,5,8}, j = {25,35}, l = {Low,Medium}, k = {1,2,3,...,12}

Suppose given no surfactant presence, and high expiratory resistance (intercept), Target Vt Kg 10, PEEP 5, and PIP 35 are present:

Time To Minimal Resistance = 81.61 +21.27 - 17.30 - 42.70 =42.88 seconds

Suppose the PEEP 0, Target Vt kg 8, no surfactant presence, high expiratory resistance (intercept), and PIP 35 are present:

Time To Minimal Resistance = 81.61 - 17.30=64.31 seconds

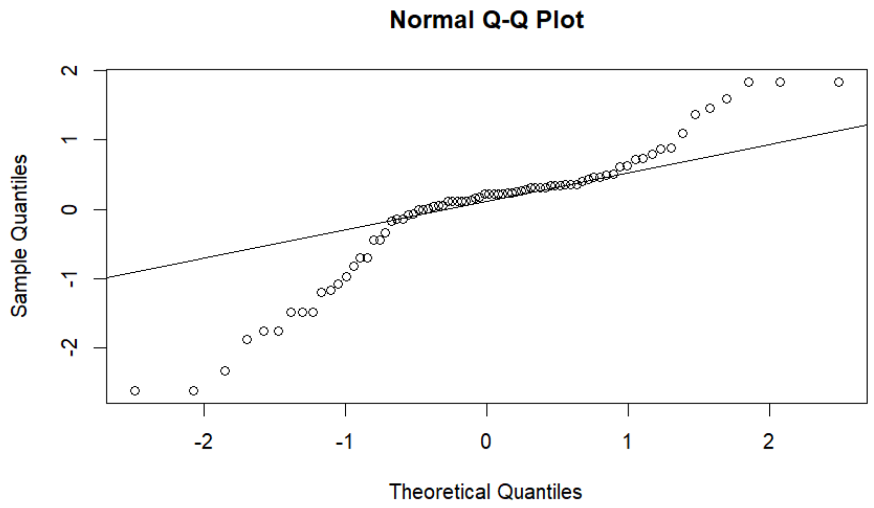

Figure 9 The QQplot tests for the normality assumption of the residuals related to the LMM. It is clear that for the most part, in the middle part, the residuals were normally distributed, except at the tails where they violate the normality assumption. Due to the low number of data points at the tails, we conclude that the normality assumption is not violated.

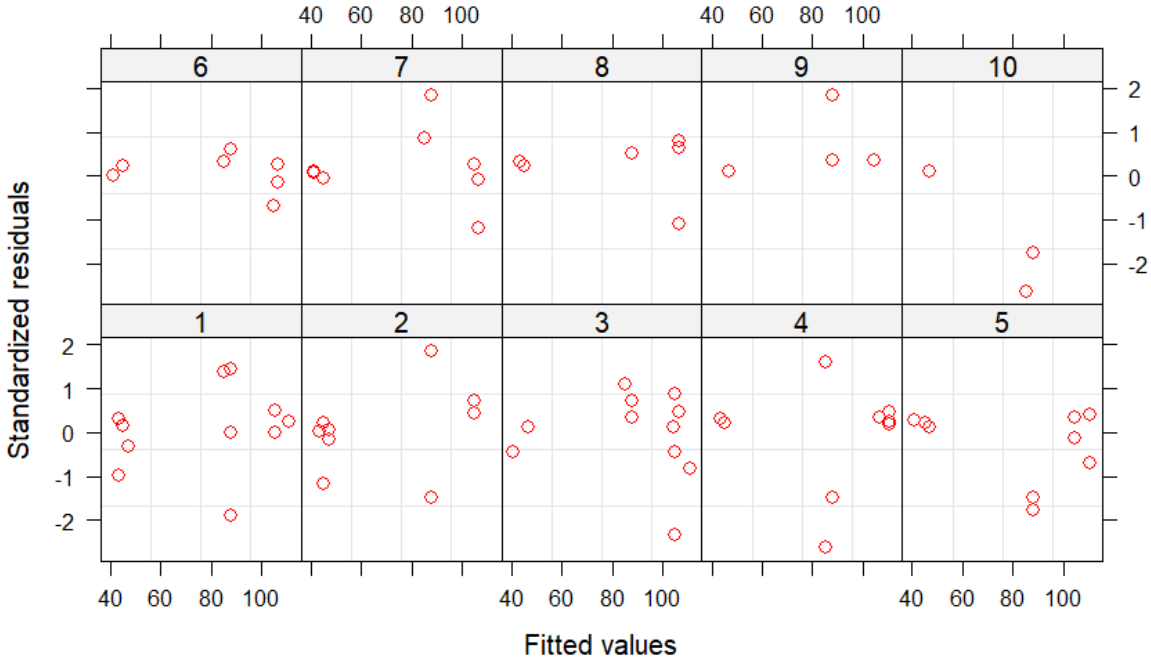

Figure 10 Test for homogeneity of the LMM. Looking at the standardized residuals of the random effect on the fitted data. Another assumption is that the standardized residuals are normally distributed and look for outliers. It can be observed that the residuals per kitten sibling group have a small spread around 0 indicating that there is no violation of the normality assumption. Outliers are within reason and do not indicate any trend or pattern.

Table 6 Overview of random effect for each sibling group. As we can see the random effect does not produce a significant effect, please note e-08 is 10^(-8).

Sibling Group	Coefficient
1	6.50e-08
2	3.58e-08
3	4.23e-09
4	-5.77e-08
5	-2.09e-07
6	4.27e-08
7	1.38e-07
8	1.06e-07
9	2.02e-07
10	-3.27e-07