# Predicting the outcome of IVF treatments using forward selection regression and linear discriminant analysis

Agnes Ekman & Linnéa Fahlberg

Fall semester 2021

## 1 Abstract

As more and more In Vitro Fertilizaiton (IVF) treatments are performed each year, there is a need to better predict the outcomes of different stages of the treatment and hence get a better understanding of which hormonal and physical parameters affect the treatment outcome and in what way. In this study, the effect of interaction between baseline AMH (anti-mullerian hormone) and DFI (DNA fragmentation index) on the chance of obtaining at least one good quality embryo was investigated, but no significance was found. Then, a statistical approach was used to find predictive models for each stage of the treatment. Linear regressions were fitted to predict continuous target variables and linear discriminant analysis (LDA) was performed to predict the binary ones. It was found that baseline AMH (anti-mullerian hormone), baseline FSH (follicle stimulating hormone), and female age significantly affect the number of retrieved oocytes (commonly referred to as eggs). Further, high BMI (body-mass index) was shown to have a significant negative impact on fertilization rate and the chance of receiving at least one good quality embryo. Finally, it was shown that the number of collected oocytes has a significant impact on fertilization rate, and that fertilization rate has a significant impact on the chance of receiving at least one good quality embryo. The best models found for predicting pregnancy and live birth did not significantly outperform the naive models which they were compared to. Hence, no significant conclusions were drawn from these models. All patients in the data set went through their first IVF treatment, defined by a first and single egg retrieval. All proven significance is on a 95% level.

# Contents

## 2 Introduction

IVF stands for in vitro fertilization and it is a medical treatment for couples who have trouble conceiving naturally. In such a treatment, eggs are retrieved from the woman and fertilized with the man's sperm outside of the body. The fertilized eggs are grown into embryos outside the body for a few days before one embryo is transferred into the woman's uterus by a medical doctor. If the embryo then implants itself in the uterine wall, a pregnancy occurs [3].

In Region Skåne (Sweden), couples have the right to three IVF treatments with egg retrieval, given that it is medically motivated. Note that one egg retrieval can result in several good quality embryos which can be frozen and then transferred into the uterus for a new pregnancy attempt. The number of egg retrievals set the limit, not the embryo transfers [6].

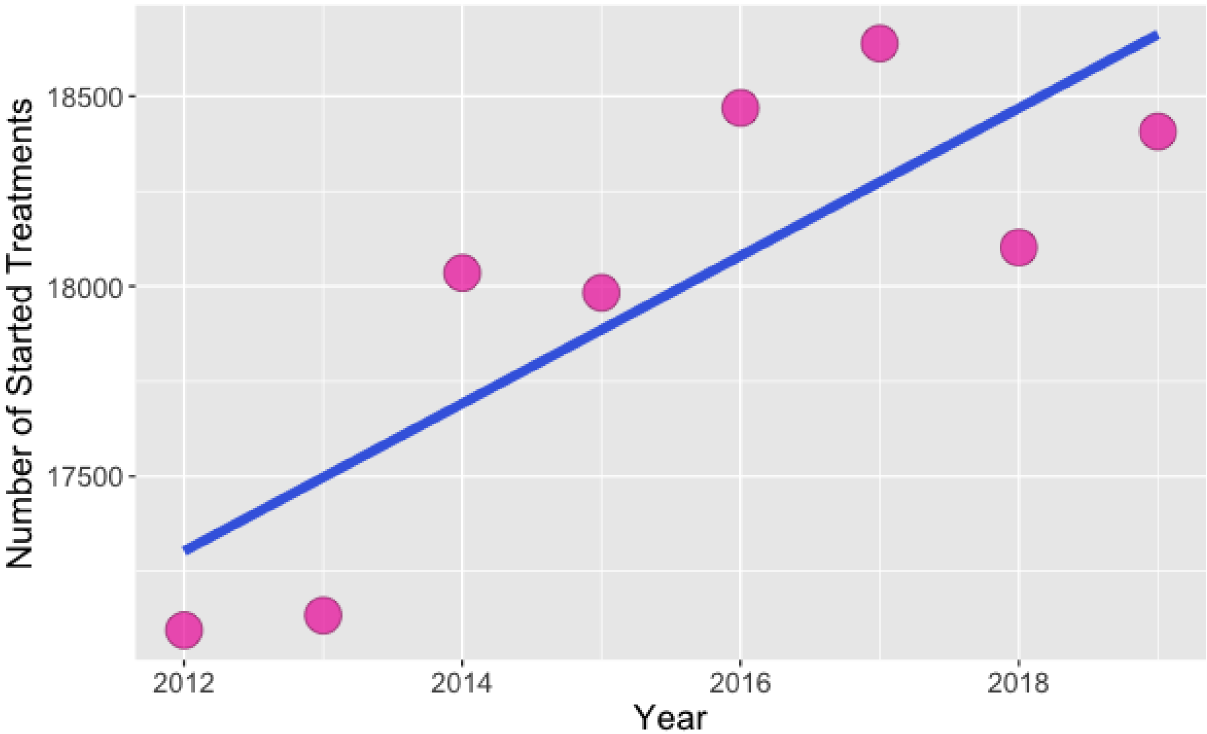


Figure 1: Number of started IVF-Treatments in Sweden each year represented by dots together with a fitted linear regression line. The number of started treatments are retrieved from the yearly reports from Q-IVF from 2014-2021.

There are two typical fertilization methods which are commonly used in IVF treatments. The first method is standard IVF, which is when the retrieved eggs are inseminated by the sperm just by mixing the two components together and letting the fertilization happen naturally. The second method is called ICSI and refers to when a single sperm is injected into the egg by needle.

IVF has had an increasing trend during the last decades, see figure 1, and are now common. In 1992, 3000 treatments were performed in Sweden and in 2020, it had increased to 22000. In total, approximately $10-15\%$ of Swedish heterosexual couples suffer from infertility [10].

## 2.1 Purpose

The purpose of this research project is to find statistical models which predict the outcome of different stages of IVF-treatments. Seven steps were identified where the first is the female start parameters and the second is the male start parameters and the choice of fertilization method. More information gets available moving on through the next steps of the treatment: Egg Retrieval, Fertilization, Embryo Transfer, and Pregnancy. The final step, and the objective of the IVF treatment, is Birth. All these stages of the treatment, as well as what information gets available in each step is exhibited in figure 2.

Naturally, all couples going through an IVF treatment is longing for a child. Hence, birthing a living child is the ultimate goal variable. However, the outcome of the birth step is highly dependent on the previous steps of the treatment. For example, a couple who do not get any fertilized eggs cannot develop any embryos, cannot get pregnant, and hence cannot birth a child. Therefore, prediction models were made for every step of the process after the two starting points. When predicting a target variable, naturally, only the parameters that are known before the stage of the target variable are allowed in the model.

In the parameter boxes in the timeline in figure 2, the target variables are marked in bold. These five variables: oocytes, fertilization rate, good quality embryos, pregnancy, and live birth, are the variables that will be predicted using statistical models in this study.

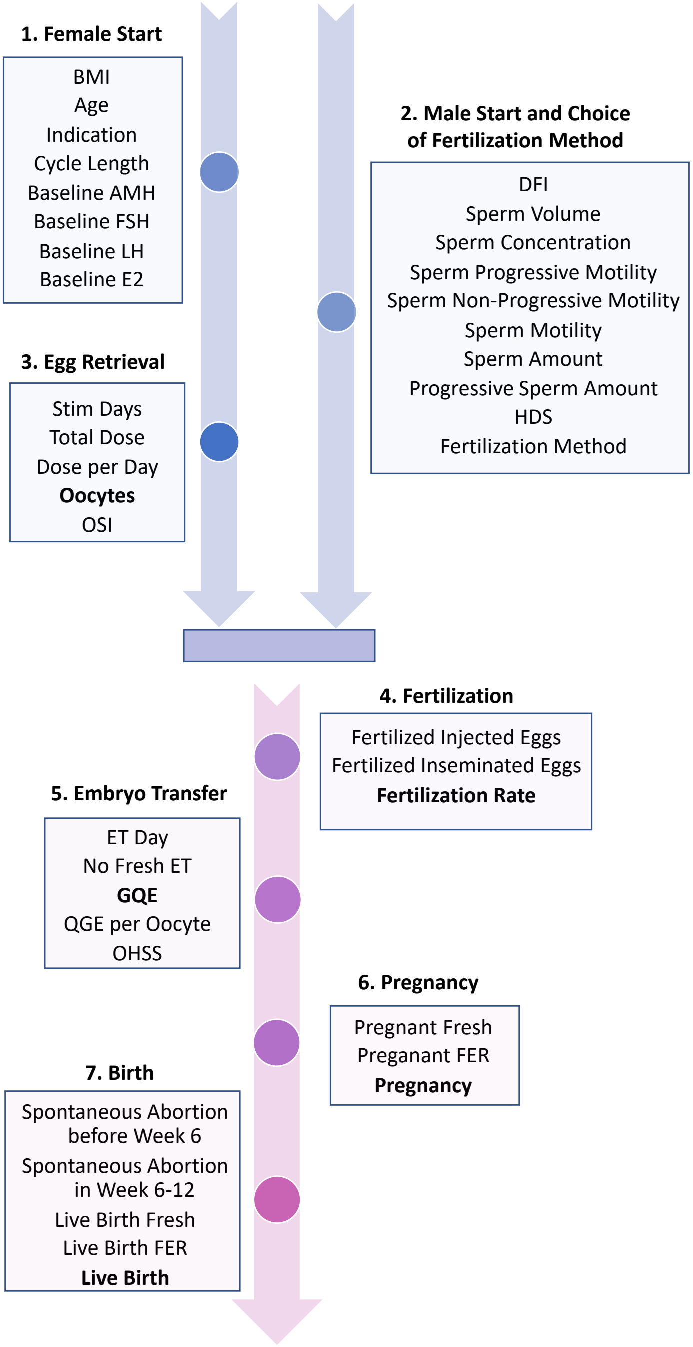


Figure 2: Parameter timeline exhibiting the different stages of the IVF treatment and how the parameters relate to them. Each treatment stage is assigned a number which will be referred to throughout the report. In the parameter boxes, the parameter marked in bold is the target variable for that stage of the treatment. All parameters are explained in the Parameter Lexicon which is found in the appendix in section 17.1.

# 3 Background

A great deal of research has been made on IVF treatments in the past. In the most common approach, researchers have looked at one single parameter at a time or at two parameters and their interaction. They have estimated a linear or logistic regression, depending on the characteristics of the target variables, and evaluated the significance of the parameters using the p-value of the regression coefficients. In most research, the data set was not divided into modelling set, validation set, and test set, but the regression was fitted to the full data set.

In this section, some previous research will be touched upon to get a sense of which conclusions have been drawn and which correlations have been found. In the discussion in section 14, the results of this research paper will be compared to and put into the context of the research which is presented below.

First, a study looking at the female anti-mullerian hormone (AMH) and the sperm DNA fragmentation index (DFI) will be introduced. This is highly relevant for this research project since it was done using the same data. Then, previous research on how body mass index (BMI) and age affect different stages of the IVF treatment will be presented. The correlation between follicle-stimulating hormone (FSH), which stimulates the egg growth in the women, is then touched upon. Finally, research on predicting the outcome of later stages of the treatment based on early results is introduced.

## 3.1 Anti-Mullerian Hormone (AMH) and DNA Fragmentation Index (DFI)

A previous study on the topic which has been made using the same data set which was used for this project was presented in [18]. The purpose of the study was to investigate the effect of the combined levels of female hormone Anti-Mullerian Hormone (AMH) and the DNA fragmentation index (DFI) as these have been shown to be individually important in the prediction of the outcome of Standard IVF. Logistic regression models were made for predicting the four binary outcome variables: ”Obtaining at least one good quality embryo”, “Pregnancy in procedures where embryo transfer was performed”, “Live Birth in all procedures”, and for “Miscarriages”.

AMH, DFI, and the interaction between the two were all proven to be statistically significant (p=0.036, p=0.001, and p=0.017) for obtaining at least one good quality embryo in standard IVF procedures. No significance of the interaction was observed for ICSI procedures. No significance was achieved when predicting pregnancy, live birth, or miscarriages.

The results of the study show that when predicting good quality embryos the marginal effect of increasing DFI by 10 percentage points was significant only when the levels of AMH were low (AMH $<$ 25.2 pmol/L). The main conclusion of the study was therefore that higher levels of DFI in the male partner can have a significant negative effect on a Standard IVF treatment, especially if the female partner has a low level of AMH [18].

## 3.2 Age and BMI

The effects of age on the number of retrieved oocytes and on the embryo quality were examined in study [7]. The authors found that both the number of retrieved oocytes and the embryo quality decreased with age.

In [14], the effects of BMI, age, and the interaction between the two were examined. The main objective was to examine the effect of BMI on the outcome of IVF treatments, but age and an interaction term was also included in the models. Linear and logistic regression were fitted, depending on the characteristic of the target variables. The results showed that higher BMI significantly affects the probability of pregnancy, and that age and the interaction between the two significantly affects the number of retrieved oocytes, the fertilization rate, the probability of pregnancy, and the probability of live birth. It was found that all fertility measures decreased with age. When it comes to the interaction between age and BMI, it was stated that BMI showed a clear negative impact on fertility for younger women, while the effect was not as pronounced for older women.

Some studies show that higher BMI is correlated with lower quality of the embryos. In study [4], it was found that BMI had a significant negative impact on the embryo quality. Another study showing the same relationship was [9], where they found that women younger than 35 years with higher BMI exhibited lower embryo quality than those with lower BMI.

Some other studies showed no significant correlation between BMI and embryo quality. One example is [17] where they found that BMI did not affect the embryo development. Another study which found that BMI did not have a significant impact on embryo quality was [16]. They intervened in the lifestyle of obese women for them to lose weight before the IVF treatment, but no increase in embryo quality was found.

When it comes to fertilization rate, BMI exhibited a significant negative impact according to [15]. In study [12], it was also found that higher BMI is correlated with lower fertilization rate. They also confirmed the result from [14], that high BMI has a negative impact on pregnancy rate. In study [5], they found no significant correlation between BMI and fertilization rate. However, they too found a significant impact on the pregnancy rate and they also found a significant negative impact of high BMI on live birth rates.

### 3.3 Baseline AMH and FSH (anti-mullerian hormone and follicle stimulating hormone)

Both baseline AMH (anti-mullerian hormone) and FSH (follicle stimulating hormone) have been shown to have a significant impact on IVF outcome. In [8], it was shown that higher levels of AMH had a significant positive effect on the number of collected oocytes, and in [1] where the objective was to investigate the reasons why women with elevated baseline FSH levels have reduced pregnancy and live birth rates, they found that high baseline levels of FSH were significantly correlated with lower number of collected oocytes.

### 3.4 Number of retrieved oocytes as predictor

In [13], the predictive value of the number of oocytes was investigated. The authors found no significant difference in the number of oocytes between women who got pregnant and those who did not.

## 4 Data

### 4.1 Subsets of the data

The data set consists of 456 couples who has gone through a first round of IVF-treatment, meaning that only one egg retrieval was performed. The data was initially divided into a modelling set (N=300) and a test set (N=156). The test set was put aside and was not looked at during the modelling stage of the project. Note that this was done before any exclusion based on criteria or missing values.

The modelling set was further divided into a training set and a validation set of size 2/3 and 1/3 respectively, relative to the data amount after removing samples based on the exclusion criteria and missing values. The model was fitted to the training set and then predictions were made on the validation set.

For all models, a bootstrap approach was used to test the significance of the predictions' performances on the validation data. Hence, every model was tested on 1000 test and validation set pairs, each randomly generated from the modelling set.

## 4.2 Exclusion criteria

Three exclusion criteria were used throughout this project. First, all couples where no insemination was performed were excluded from the data set (N=10). This was defined as the fertilization method or the sum of injected and inseminated eggs being equal to zero. Second, all couples with a fertilization rate larger than one were excluded (N=2). Finally, all couples that became pregnant without any successful fertilization were excluded (N=5). These three sets did not overlap, excluding a total of 17 samples and leaving 439 samples in the data set.

In each step, those treatments that already failed were excluded from the data set. This was done because it is certain that a treatment which has already failed will exhibit negative outcomes in all following stages. Hence, no value is added by predicting the outcome using a stochastic model, assuming that all information available before the target variable is known at the moment of the prediction. For example, the outcome of the pregnancy stage is assumed to be known when predicting live birth, and hence, if there is no pregnancy then it is certain that there will be no live birth. No statistical model is needed to draw that conclusion and therefore those samples are excluded.

Another approach which is commonly used in medical research is to predict all target variables using only the start parameters which are known from the beginning of the treatment, stage 1 and 2 in the timeline in figure 2. Note that this is not how this study was formed. Hence, when interpreting the results of this study, it is important to bear in mind that the models are created using all available information up until the predicted stage of the treatment.

The first step to be modelled was step 3, where the number of retrieved oocytes was the target variable. At least one oocyte (step 3) was retrieved from each couple. Hence, all data points were used when modelling fertilization rate (step 4). However, when predicting whether or not the couple would get at least one good quality embryo (step 5), all couples that had a fertilization rate of zero were excluded (N=26). Moving on to predicting pregnancy (step 6), all couples that had zero good quality embryos were excluded (N=32). Finally, when modelling live birth (step 7), all couples that did not get pregnant were excluded (N=172).

## 4.3 Missing values

In all models, only complete data points were used. Those data points that had missing values for any model parameter were excluded for that particular model. As a result, the data set varies slightly throughout the report, depending on which parameters are included in the current model.

The reason why this approach was used is because a complete set is needed for the models to function properly, but removing all data points that had any missing value would result in a too large reduction of the data set. Hence this approach, only excluding those points that have missing values for the relevant parameters for each model, made it possible to always use as many data points as possible. This is desirable since the more data points that are used for the parameters estimates, the lower is the estimates' variances.

One drawback of this approach is that the difference in the data sets gives the models slightly different conditions to form high performing predictions.

An alternative approach is imputation of missing values. Because of the small size of the data set and high uncertainty of the parameters, this approach did not seem effective in this case.

## 4.4 Handling unbalance with oversampling

When predicting whether or not there would be at least one good quality embryo and whether or not there would be a live birth, the data sets are problematically unbalanced when it comes to the outcome. 92% of the couples in the GQE data set (after exclusion based on criteria) got at least one good quality embryo, and 76% of the couples in the LB data set birthed a living child.

The unbalance causes problems since it encourages the model to only predict the positive outcomes, since this would result in high accuracy of 92% and 76%, respectively. However, this approach yields a specificity of 0 and does not add any predictive value when it comes to separating the two groups. A good predictor correctly predicts both the positive and negative outcomes.

To handle this problem, oversampling was used in the training set, which means that randomly chosen data points from the smaller group were duplicated to match the data amount of the bigger group. Hence, the smaller group will be given as much weight as the larger and the model will try to predict them with equal accuracy.

One problem that persists is that the small data amounts of the smaller groups in the validation sets yields a high uncertainty in the specificity. This is because testing on few data points makes the result more dependent on the random choice of the validation points, than if a large validation set was available where most of the characteristics would be represented. Hence, the significance of the specificity is remarkably low for these models. To solve this issue, a larger or more balanced data set is needed to decrease the uncertainty.

# 5 Method

## 5.1 Target variables

For each step of the IVF treatment, from egg retrieval to birth, a target variable was chosen for the prediction models. The variables used as targets were oocytes (step 3), fertilization rate (step 4), good quality embryos (step 5), pregnancy (step 6), and live birth (step 7). See the timeline in figure 2 for an overview of the process, its steps and its target variables.

The target variables were either seen as a continuous variable or a binary one. For the binary target variables, a positive outcome was defined as 1 and a negative as 0. Linear regressions were fitted to predict the continuous target variables, and linear discriminant analysis were performed to predict the binary.

## 5.2 Linear regressions

When modeling data in a multiple linear regression it is assumed that, on average, the target variable, Y, has a linear relationship with the variables $X_1$,...$X_p$ and follows the following linear model:

$$E[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 ... + \beta_p X_p \tag{1}$$

Were $\beta_0$ is the model intercept and $\beta_i$, the parameter coefficient for $X_i$, is the slope in the $X_i$ direction.

For a single observation the model can be expressed as:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} ... + \beta_p X_{ip} + \epsilon_i \tag{2}$$

Were:

$Y_i$, is the ith measurement of the target variable for i=1,...,n, were n is the total number of observations of Y.
$X_ij$ is the ith observation of variable $X_j$ for j=1,...,p.
$\beta_0$ is the model intercept.
$\beta_i$ is the parameter coefficient for $X_i$.
$\epsilon_i$ is the measurement error.

When data is fitted into a multiple linear regression model, the parameter coefficients are calculated so the measurement errors are minimized.

For this study, when fitting regressions, a null model containing only an intercept and a full model containing all parameters known before the step of the target variable were defined. A forward selection using the Akaike information criterion (AIC) and one using the Bayesian information criterion (BIC) were performed, resulting in two model candidates. The BIC criterion favours smaller models, and will hence most likely contain fewer parameters.

The two models were compared using the mean absolute error (MAE), defined as:

$$\text{MAE} = \frac{\sum_{i=1}^{n} |\epsilon_i|}{n} \tag{3}$$

When the MAE:s were not significantly different, the BIC model was chosen because of the lower risk of overfitting due to the number of parameters.

The final model was compared to a naive model, only containing an intercept, to investigate the significance of the additional predictive value of the model. The significance of the parameter coefficients was also checked, to make sure that only parameters with coefficients significantly different from zero were included in the models.

## 5.3 Linear discriminant analysis, LDA

Linear discriminant analysis (LDA) is a classification method and it can therefor be used to predict binary target variables. LDA models projects the samples onto a line which maximizes the separation between the two groups of the target variables. When performing an LDA, observations are given an LDA-score, Z, which is a linear combination of variables. The LDA-score is defined as:

$$Z = \beta_1 X_1 + \beta_2 X_2 ... + \beta_p X_p \tag{4}$$

Where Z is the LDA-score given to an observation. Observations with an LDA score below zero are classified with a negative outcome whereas observations with an LDA score of zero or above are classified with a positive outcome. $X_i$ are normalized variables used as predictors in the model. The $\beta$-coefficients are normalized model coefficients which are calculated so the separation between groups is maximized.

For the linear discriminant analysis in this study, a full model was defined which included all parameters known before the step of the target variable. An LDA model was fitted to the full model. As the $\beta$-coefficients of the LDA model are normalized they represent the importance of the parameters when it comes to explaining the variability of the target variable. The estimated $\beta$-coefficients of the LDA models are hence referred to as the weights of the parameters.

The normalized parameter weights were then used to form reduced models. First, only the two parameters with the highest weights were added. Then the parameters which had weights larger than 0.05 were added one at a time. If the addition of a parameter increased the performance of the model the parameter was kept, but if it did not increase the performance it was excluded since a larger model comes with a higher risk of overfitting. When all parameters with a weight larger than 0.05 which increased the performance were added, all parameters in the model were excluded one at a time, starting with the parameter that was added first, to check if they still had a positive effect on the performance.

Four types of performance measures were used to compare the models: accuracy, F1-score, sensitivity, and specificity. All predictions made can be put in either one of the four following categories: True positives (TP), the number of observations which have been correctly classified with a positive outcome, true negatives (TN), the number of observations which have been correctly classified with a negative outcome, false positives (FP), the number of observations which have been incorrectly classified as positive, and false negatives (FN), the number of observations which have been incorrectly classified as negative. These help define the performance measures used in this study.

Accuracy, the ratio of predictions which have been correctly classified, is defined as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5}$$

Sensitivity, the ratio of predictions which have been correctly classified as positive, is defined as:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

Specificity, the negative equivalent of sensitivity or the ratio of predictions which have been correctly classified as negative, is defined as:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{7}$$

Higher sensitivity often comes at the expense of the specificity, and vice versa. The F1-score is a metric which favours models which can correctly classify both positive and negative outcomes. It is defined by both the sensitivity and a measurement called precision which is defined as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{8}$$

Finally, the F1-score is defined as:

$$\text{F1} = 2 \times \frac{\text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}} \tag{9}$$

The performance was deemed to increase if neither accuracy nor F1 decreased, and at least one of them increased with at least 0.3 percentage points. Three exceptions were made when the inclusion or exclusion of the parameter had a substantial effect on the specificity in the opposite direction of the accuracy and F1. These exceptions had little to no effect on the final models after all selection was done.

The final model was compared to a naive, ”flip a coin” model, which randomly generates 0 and 1 at a probability of 0.5, to investigate the significance of the additional predictive value of the model. The significance of the parameter coefficients was also checked, to make sure that only parameters with coefficients significantly different from zero were included in the models.

# 6 Hypothesis tests

When testing the significance of the parameter estimates and performance of the final models, 95% confidence intervals were used. The parameters were deemed significant if their two-sided confidence intervals did not cover zero.

The performance of the final models were compared to that of naive models, as described above. Both the naive model and the final model were estimated and validated in each step of the bootstrap loop. Hence, paired samples were used to test significance, where the performance of the naive model was subtracted from that of the final to compute the test statistics.

Since the final model is expected to perform better than the naive model, significance was achieved in the LDA models when the lower limited one-sided confidence interval did not cover zero. This is because for all performance measures used for the LDA model, a higher number indicates better performance. In the linear regression case, the relationship is opposite since a lower MAE (Mean Absolute Error) indicates better performance. Hence, upper limited confidence intervals were used to test significance for the linear regression models' performance.

# 7 Revisiting interaction between AMH and DFI

In this section, the previous study [18] made on the same data set was revisited. The same method was followed as in the previous research, fitting a logistic regression with DFI, Baseline AMH, and the interaction between the two, modelling on all data points in the full data set and using the p-value to test significance. Since obtaining at least one good quality embryo was the only target variable where significance was found for the interaction between AMH and DFI, this is the target variable which was revisited.

The same exclusion criterias were used as in the previous study [18]: established

PCOS diagnosis (N=18(25)), the use of both standard IVF and ICSI procedures on retrieved oocytes (N=38), Non-ejaculated spermatoza (N=26(32)), no attempted inseminations (N=10). Note that there was a difference in the number of patients with a PCOS diagnosis, 18 in this data set versus 25 in the previous study, and with non-ejaculated sperm, 26 versus 32. Since the exclusion criteria overlap, this resulted in a total of 11 couples being included here which were excluded in the previous study. This difference could be due to some retroactive changes in the data.

There is one major difference in the data set which likely affects the results: the DFI values which were missing (N=76) at the time of Zarén et al's study has now been added to the data set. They used imputation to estimate the missing values, while now the true values are available for all patients.

When fitting a logistic regression to the current data set, containing all the true values for DFI, only DFI turned out to be significant (p=0.013). DFI was also deemed to be significant in [18]. Baseline AMH was not significant in [18], and was not now with the true DFI values either (p=0.456). Finally, the interaction between DFI and Baseline AMH was significant in [18] but was not found to be so when the imputed DFI values were substituted by the measured (p=0.125). See the coefficient estimates with their confidence interval and p-values in table 1.

| Parameter | Estimate | Conf. int. (95%) | p-value |
|---|---|---|---|
| Baseline AMH | -0.013 | [-0.046 0.024] | 0.456 |
| DFI | -0.076 | [-0.139 -0.017] | 0.013 |
| DFI:Baseline AMH | 0.002 | [-0.001 0.005] | 0.125 |

Table 1: Revisiting interaction between baseline AMH and DFI. Coefficient estimates of the logistic regression with their corresponding 95% confidence intervals and p-values. Colon marks the interaction term.

# 8 Prediction of oocytes

The first step of the treatment to predict is the egg retrieval (step 3), where the target variable is the number of retrieved oocytes. The oocytes parameter takes on discrete numbers ranging from 1 to 33. Hence, fitting a linear regression seems like an appropriate approach.

To choose which parameters to include in the model, forward selection was used. The idea was to add parameters one at a time, always adding the one that would increase the performance the most based on a specified criterion. The procedure was repeated until no new parameter could increase the performance.

As mentioned, the AIC criterion was used first, where a lower value indicates a better model. The biggest drawback of the AIC is that it generally favours larger models, since these by default have a better chance at fitting the values in the training set. However, larger models are harder to work with and come with a higher risk of overfitting. Hence, a second forward selection was performed with BIC as criterion. The BIC metric benefits smaller models.

The resulting two models, using the two different criteria, were then used to predict the number of fertilized eggs in the validation set. Finally, the mean absolute error was used to determine which model performed the best.

## 8.1 Forward selection AIC

| Step | Parameter | Estimate | Conf. int. (95%) |
|---|---|---|---|
| 0 | Intercept | 1.548 | [-3.139 6.456] |
| 1 | Baseline AMH | 0.121 | [0.090 0.156] |
| 2 | Age | 0.222 | [0.082 0.354] |
| 3 | Baseline FSH | -0.342 | [-0.575 -0.117] |
| 4a | Indication 2 | -1.550 | [-3.039 -0.044] |
| 4b | Indication 3 | 0.268 | [-0.819 1.286] |
| 4c | Indication 4 | -1.391 | [-3.199 0.476] |
| 5 | Baseline E2 | 0.002 | [0.000(+) 0.004] |
| 6 | Baseline LH | -0.022 | [-0.076 0.031] |

Table 2: Forward selection using AIC criterion to find a linear regression model for predicting the number of retrieved oocytes. The forward selection steps are numbered, and the corresponding parameter estimates are exhibited with their corresponding 95% confidence interval.

Starting with a null model with only an intercept, the AIC forward selection added parameters according to the steps exhibited in table 2. Baseline AMH (anti-mullerian hormone) was added first, indicating that this is the parameter which explains most of the variability of the number of retrieved oocytes. Its coefficient was significant, since its confidence interval [0.090 0.156] does not cover zero. Since the coefficient is estimated as positive, a higher level of baseline AMH is correlated with a higher number of retrieved oocytes.

The next parameters to be added by the AIC criterion were age and baseline FSH (follicle-stimulating hormone). These two were both also significant with confidence intervals which do not cover zero, [0.082 0.354] and [-0.575 -0.117] respectively. Since age has a positive coefficient, it is more likely to retrieve many oocytes from older women than from younger according to the model. The coefficient for baseline FSH is negative, implying a negative correlation between the parameter and the target variable, meaning that a higher baseline FSH level results in fewer retrieved oocytes.

Indication, which explains the reason behind the infertility, was added next as step 4 of the forward selection. Indication 2, 3, and 4, are exhibited in table 2 as 4a, b, and c. Indication 2 (female factor) was significant, with a confidence interval which does not cover zero ([-3.039 -0.044]). Indication 3 and 4 however did not exhibit significance, with confidence intervals [-0.819 1.286] and [-3.199 0.476] which both cover zero.

In step 5, baseline E2 was added, which is another female hormone. The coefficient was (barely) significant and positive, indicating that a higher E2 level is correlated with more oocytes.

Finally, baseline LH (luteinizing hormone) was added in step 6. This coefficient was however not significant, with a confidence interval that covers zero [-0.076 0.031].

To stick with the AIC criterion, the insignificant parameters, Indication 3, Indication 4, and Baseline LH, were all included when computing the mean absolute error (MAE) of 4.215 [3.610 4.833] and comparing it to that of the BIC, which will be introduced shortly. Note however that these parameters are insignificant at a 95% confidence level, indicating a high uncertainty and risk of overfitting.

The AIC value after each step is exhibited in the appendix in table 32. There, it is shown how the AIC drops for each added variable as it should, and that the decreases are smaller for the later steps than for the first ones, indicating that the parameters which were added last do not increase the performance as much as the first ones.

To summarize, the optimal model based on forward selection using the AIC criterion exhibited a MAE estimate of 4.215 and included the seven parameters exhibited in table 2: baseline AMH, age, baseline FSH, indication, baseline E2, and baseline LH.

## 8.2 Forward selection BIC

| Step | Parameter | Estimate | Conf. int. (95%) |
|---|---|---|---|
| 0 | Intercept | 2.596 | [-2.405 7.048] |
| 1 | Baseline AMH | 0.119 | [0.088 0.154] |
| 2 | Age | 0.221 | [0.096 0.362] |
| 3 | Baseline FSH | -0.395 | [-0.625 -0.158] |

Table 3: Forward selection using BIC criterion to find a linear regression model for predicting the number of retrieved oocytes. The forward selection steps are numbered, and the corresponding parameter estimates are exhibited with their corresponding 95% confidence interval.

When using BIC as criterion instead, the forward selection took the same steps as the initial ones when using the AIC, but now it stopped after only including the three parameters exhibited in table 3: baseline AMH, age, and baseline FSH. As expected, this model is smaller than the one selected using the AIC criterion since the BIC metric favours small models.

In this model too, the coefficients for baseline AMH, age, and baseline FSH are all significant with confidence intervals [0.088 0.154], [0.096 0.362], and [-0.625 -0.158] respectively. Again, AMH and age exhibit positive correlation with the target variable while the correlation with FSH is negative.

The BIC values of each step of the forward selection is exhibited in the appendix in table 33. The MAE of the optimal BIC was estimated to 4.114 with the confidence interval [3.547 4.690].

## 8.3 Model Comparison

| Model | Mean absolute error | Conf. int. (95%) |
|---|---|---|
| AIC model | 4.215 | [3.610 4.833] |
| BIC model | 4.114 | [3.547 4.690] |

Table 4: Mean absolute error on validation set with the AIC and BIC models from forward selection when performing linear regression with oocytes as target Variable.

The BIC model has a mean absolute error of 4.11 [3.55, 4.69], which is smaller than that of the AIC model which was 4.21 [3.61, 4.83]. Since BIC is also smaller with fewer parameters, it is clearly the better model of the two.

A hypothesis test was performed to try if the BIC model is significantly better than a naive model which only estimates an intercept. A one-sided test was formed as:

$$\begin{aligned} H0 :& MAE_{BIC} - MAE_{Naive} = 0 \\ H1 :& MAE_{BIC} - MAE_{Naive} < 0 \end{aligned} \tag{10}$$

The confidence interval of the test statistic, $MAE_{BIC} - MAE_{Naive}$, was $[-\infty, -0.177]$. Since it does not cover zero, the BIC model has a significantly lower MAE and hence performs significantly better than the naive model. The test statistic is visualized as a histogram in figure 3. It looks somewhat normally distributed. The pink area represents the 95% upper limited confidence interval, and it is clear that it does not cover zero.

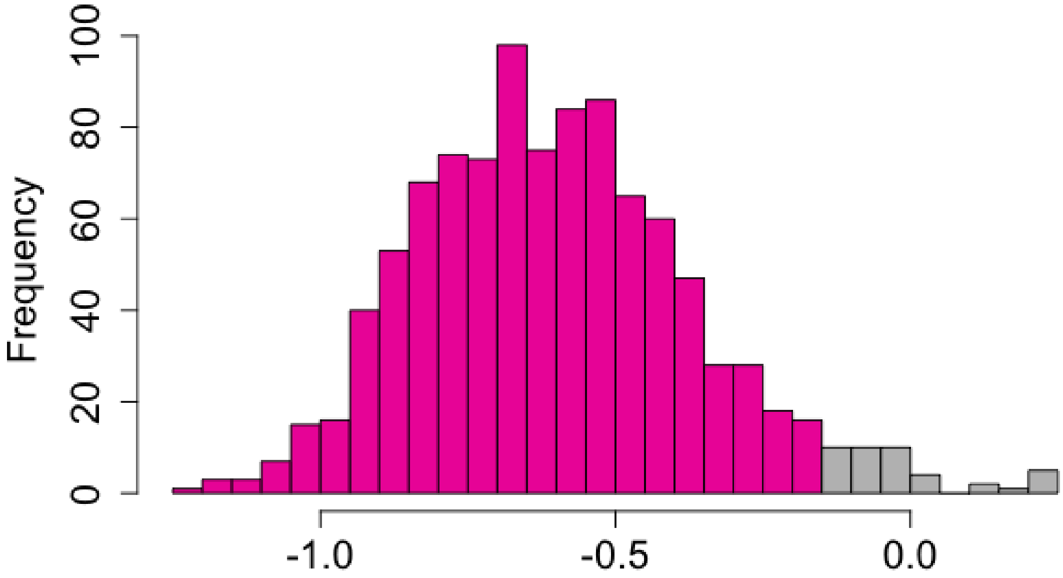


Figure 3: Histogram of mean absolute errors for the BIC model pairwise subtracted by that of the naive model, when predicting the number of retrieved oocytes. The errors are retrieved from the bootstrap loop.

# 9 Prediction of fertilization rate

When modelling the next step of the treatment, fertilization (step 4), the target variable was fertilization rate. It is a continuous variable on the interval [0 1]. As a first attempt, a linear regression was fitted using the same method as in section 8. However, a linear regression does not take into account that the target variable is limited to a closed interval.

## 9.1 Linear regression

### 9.1.1 Forward selection AIC

| Step | Parameter | Estimate | Conf. int. (95%) |
|---|---|---|---|
| 0 | Intercept | 0.367 | [0.150 0.586] |
| 1 | Age | 0.005 | [-0.001 0.011] |
| 2 | Sperm Concentration | 0.001 | [0.000(+) 0.001] |
| 3a | Indication 2 | -0.007 | [-0.106 0.094] |
| 3b | Indication 3 | 0.066 | [0.011 0.118] |
| 3c | Indication 4 | -0.095 | [-0.178 -0.014] |
| 4 | Oocytes | -0.009 | [-0.016 -0.003] |
| 5 | OSI | 0.007 | [0.002 0.016] |

Table 5: Forward selection using AIC criterion to find a linear regression model for predicting the fertilization rate. The forward selection steps are numbered, and the corresponding parameter estimates are exhibited with their corresponding 95% confidence interval.

As in section 8, a forward selection was performed using the AIC as criterion. The steps are exhibited in table 5. The first parameter to be added, hence explaining the most of the variability in the target variable, was age. However, the coefficient was not significant, with a confidence interval of [-0.001 0.011] which covers zero. This indicates that there is a high level of uncertainty in the model.

Second, sperm concentration was added to the model. The coefficient was significant with confidence interval which does not cover zero, [0.000(+) 0.001], exhibiting a positive correlation between sperm concentration and fertilization rate.

Indication was added next, shown as step 3a, b, and c, in table 5. Indication 3 (male factor) was significant with confidence interval [0.011 0.118], indicating a positive correlation between indication 3 and fertilization rate. Indication 4 (multiple factors) was also significant with confidence interval [-0.178 -0.014]. The negative coefficient estimate points at a negative correlation with the target variable. The last indicaiton, Indication 2 (female factor), however was not significant since its confidence interval [-0.106 0.094] covers zero.

The number of oocytes was added in the fourth step of the forward selection. Its coefficient was estimated to -0.009 with the confidence interval [-0.016 -0.003], exhibiting a significant negative correlation with the fertilization rate. This leads to the counter intuitive conclusion that more retrieved eggs leads to a lower fertilization rate.

Finally, OSI (Ovarian Sensitivity Index) was added in a fifth step of the forward selection. Its coefficient was significant with a confidence interval which does not cover zero, [0.002 0.016]. The estimated correlation with the fertilization rate was positive, meaning that a higher OSI leads to a higher fertilization rate.

As in section 8, even the insignificant parameters, age and indication 2, were included when computing the MAE and comparing it to the BIC model. However, it should be emphasized that these parameters are not significantly correlated with fertilization rate on a 95% confidence level.

To summarize, the optimal model contained the five parameters in table 5: age, sperm concentration, indication, oocytes, and OSI. Sperm concentration, indication 3, oocytes, and OSI were all significant since their confidence intervals do not cover zero. Age and indication 2 and 4, however, had confidence intervals that covered zero and are hence insignificant. The final model exhibited a MAE of 0.219 with the confidence interval [0.192 0.246].

### 9.1.2 Forward selection BIC

When the forward selection was repeated using BIC as criterion instead of AIC, no parameters were added and the model chosen was the null model containing only an intercept, see table 6. This again points at the uncertainty of the model found by AIC in section 9.1.1.

| Step | Parameter | Estimate | Conf. int. (95%) |
|---|---|---|---|
| 0 | Intercept | 0.517 | [0.496 0.540] |

Table 6: Forward selection using BIC criterion to find a linear regression model for predicting the fertilization rate. The forward selection steps are numbered, and the corresponding parameter estimates are exhibited with their corresponding 95% confidence interval.

### 9.1.3 Model Comparison

| Model | Mean absolute error | Conf. int. (95%) |
|---|---|---|
| AIC model | 0.219 | [0.192 0.246] |
| Naive predictor/BIC model | 0.217 | [0.192 0.242] |

Table 7: Mean absolute error on validation set with the AIC and BIC models from forward selection when performing linear regression with fertilization rate as target variable. In this case, the BIC model is the same as the naive, only fitting an intercept.

The best BIC model, which is the naive model only containing an intercept, exhibits a MAE of 0.217 which is smaller than that of the best AIC model. Hence, the AIC model does not outperform the simpler BIC/naive model.

Even though some coefficients in the best AIC model were significantly different from zero, they did not significantly increase the precision of the predictions. Hence, the conclusion that these parameters significantly affect the fertilization rate can not be drawn.

## 9.2 LDA Classification

Since the linear regression approach did not predict the fertilization rate well, a second approach was taken where a linear discriminant analysis was performed to predict if the couple would get a fertilization rate above or below 50%, see table 8. 50% was chosen both because it is an intuitive border and because it coincides well with the sample mean. After categorizing the fertilization rate as explained by table 8 and thus creating the target variable Fertilization group, a classification model was sought as the target variable is now a binary outcome. Hence, a linear discriminant analysis was performed.

| Fertilization group: 0 | Fertilization group: 1 |
|---|---|
| Fertilization rate $< 0.5$ | Fertilization rate $\geq 0.5$ |

Table 8: Categorization of fertilization rate into a binary variable which is used as target variable in the LDA below.

### 9.2.1 LDA All parameters

All parameters which were known before the fertilization (step 4) were included in a first LDA model, see timeline in figure 2 for an overview of the parameters. The parameter estimates of the LDA are normalized in a way that they can be used to compare the importance of each parameter. The parameter with the largest absolute value of its estimate is the parameter which explains most of the variability of the target variable. Hence, a large model where all parameters known before the fertilization step are included gives an indication of which parameters are of highest importance.

in table 9 the order of importance of the parameters is exhibited for parameters which had an absolute weight larger than 0.05. The results shown in this table indicate that fertilization method 2 and 3 were the two most important parameters when predicting the fertilization group of the IVF-patient. Other important parameters when modeling Fertilization group are, in descending order; Indication 4, Stim Days, Indication 2, Indication 3, The number of retrieved oocytes (Oocytes), BMI, OSI, Sperm Volume, and Age. The table only shows those parameters which had an absolute weight larger than or equal to 0.05 as those were the only parameters which were included when reduced models were created when modeling Fertilization Group. To see the entire ranking of all the parameters included in this model and their exact absolute weights, see table 36 in section 17.4 in the appendix.

| Order | Parameter |
|---|---|
| 1 | Fertilization method 2 |
| 2 | Fertilization method 3 |
| 3 | Indication 4 |
| 4 | Stim Days |
| 5 | Indication 2 |
| 6 | Indication 3 |
| 7 | Oocytes |
| 9 | BMI |
| 10 | OSI |
| 11 | Sperm Volume |
| 12 | Age |

Table 9: The parameters of the LDA model predicting fertilization group. The parameters are exhibited in descending order based on their absolute weight. Only parameters with an absolute weight $\geq 0.05$ are included since these were the ones included in the selection of a reduced model. To see the order of all parameters and their absolute weights, see table 36 in section 17.4 in the appendix.

### 9.2.2 LDA Reduced

After creating the full model, where the parameter weights rank the importance of the parameters, a reduced model using fewer parameters and performing better than the initial one was sought. The first step was to create a model with the two parameters explaining the most of the variability, Fertilization method and Indication (see table 9). Next, the parameters displayed in table 9 were systematically added to the model one at a time in order of importance. If the addition of a parameter increased the performance the parameter was kept in the model, but if it did not increase the performance it was excluded since a larger model comes with a higher risk of overfitting. To see the exact criteria used when deciding whether to keep a parameter or not in the model, see section 5.3.

When all parameters with a weight larger than 0.05 that increased the performance were added, all parameters in the model were excluded one at a time, starting with the parameter that was added first, to check if they still had a positive effect on the performance. To see each step of the process where parameters were first added systematically in order of importance and then later reduced in the same order as they were added see table 37 in the appendix. The table displays the accuracy, F1-score, sensitivity, and specificity of all the models which were tested on the way to find the best performing model.

After the selection process of adding and excluding parameters, the best performing model included fertilization method, oocytes, and BMI, and its performance measures are exhibited in table 10. The table also includes a naive predictor, which is a model predicting the fertilization group of the IVF patient by randomly generating 0 and 1 with a probability of 50% each. This model is used for comparison as a model is only deemed to add predictive value if it can significantly outperform a naive predictor.

| Model | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|
| Naive predictor | 0.50 [0.40 0.61] | 0.53 [0.42 0.65] | 0.50 [0.37 0.63] | 0.51 [0.34 0.67] |
| Fert. Method, Oocytes, BMI | 0.63 [0.55 0.71] | 0.72 [0.65 0.79] | 0.82 [0.66 0.96] | 0.35 [0.20 0.53] |

Table 10: Accuracy, F1-score, sensitivity, and specificity for the naive predictor and the optimal reduced LDA model with fertilization group as target variable. Each performance measure is presented with its mean and its 95% confidence interval.

The significance of each parameter in the best performing model, fertilization method, oocytes, and BMI, was tested using the 95% confidence intervals of the coefficients. If the 95% confidence interval of the parameter estimate covers zero the parameter is insignificant and if it does not cover zero it is significant. The results of the parameter estimation can be seen in table 11. The table shows that Fertilization method 2 (ICSI), Oocytes and BMI all exhibited significant coefficients with confidence intervals which do not cover zero: [1.185 1.978], [-0.131 -0.047], and [-0.255 -0.087] respectively. Fertilization method 3 (micro combination), however, is insignificant as its confidence interval [-0.143 2.247] covers zero.

| Parameter | Mean | Conf. int. (95%) |
|---|---|---|
| Fertilization method 2 | 1.619 | [1.185 1.978] |
| Fertilization method 3 | 1.103 | [-0.143 2.247] |
| Oocytes | -0.092 | [-0.131 -0.047] |
| BMI | -0.178 | [-0.255 -0.087] |

Table 11: Parameter estimates for the LDA model that best predicted the fertilization group, which performance was exhibited in table 10. The estimates are presented as well as their 95% confidence interval.

Since the parameter Fertilization method 3 was insignificant, a final model was created where Fertilization method 3 was excluded from the best performing model.

A dummy variable was created for fertilization method 2, to exclude the insignificant fertilization method 3 from the model. It was created as defined by table 12.

| Fertilization method 2 dummy: 0 | Fertilization method 2 dummy: 1 |
|---|---|
| Fertilization method $\neq 2$ | Fertilization method $= 2$ |

Table 12: Defining a dummy variable for fertilization method 2

The performance of the final model, including only the significant parameters Fertilization method 2 (ICSI), Oocytes, and BMI, is exhibited in table 13, where the naive predictor is included as well for comparison and so is the test statistic called paired difference, which is the performance of the final model pairwise subtracted by that of the naive model. For example, the paired difference of accuracy is the accuracy of the final model subtracted by the accuracy of the naive model.

The paired difference of the models' accuracy shows that the final model has a significantly higher accuracy since the confidence of the paired difference, [0.03 $\infty$], does not cover zero. The same goes for F1-score and sensitivity with confidence intervals [0.08 $\infty$] and [0.18 $\infty$], respectively. The final model did not significantly outperform the naive when it comes to specificity, however, since the paired difference has a confidence interval which covers zero, [-0.34 $\infty$].

| Model/<br>Comparison | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|
| Naive predictor | 0.50 [0.40 0.61] | 0.53 [0.42 0.65] | 0.50 [0.37 0.63] | 0.51 [0.34 0.67] |
| Oocytes, BMI<br>Fert. method 2 | 0.64 [0.55 0.71] | 0.73 [0.66 0.80] | 0.84 [0.68 0.96] | 0.34 [0.20 0.51] |
| Paired difference | 0.14 [0.03 $\infty$] | 0.19 [0.08 $\infty$] | 0.34 [0.18 $\infty$] | -0.16 [-0.34 $\infty$] |

Table 13: Accuracy, F1-score, sensitivity, and specificity for the naive predictor and the final LDA model, when only including significant parameters, still with fertilization group as target variable. Each performance measure is presented with its mean and its 95% confidence interval. The paired difference between the final and naive model is also included with its lower limited 95% confidence interval.

| Parameter | Mean | Conf. int. (95%) |
|---|---|---|
| Fertilization method 2 | 1.548 | [1.105 1.919] |
| Oocytes | -0.095 | [-0.136 -0.047] |
| BMI | -0.180 | [-0.265 -0.084] |

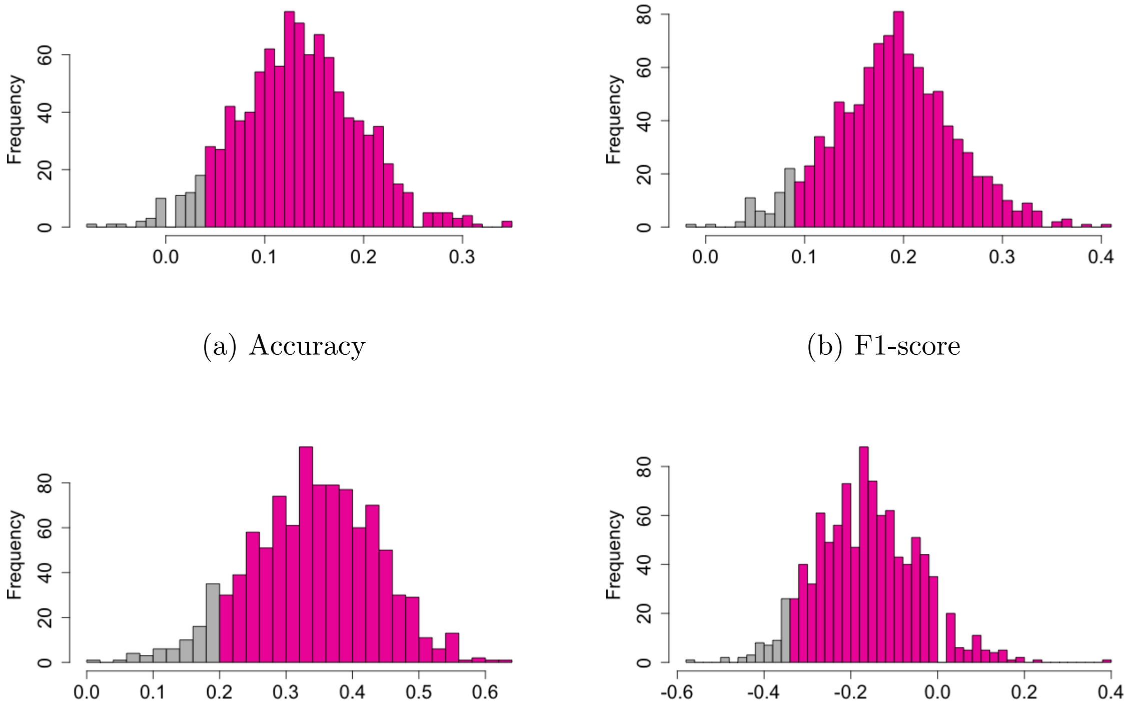
Table 14: Parameter estimates for the LDA model predicting fertilization group using only significant parameters, which performance was exhibited in table 13. The estimates are presented as well as their 95% confidence interval.

The parameters' coefficients were re-estimated for the final model, excluding the insignificant parameter Fertilization method 3, and the results of that can be seen in table 14. The model shows a positive correlation between Fertilization method 2 (ICSI) and the target variable Fertilization Group as the parameter estimate is estimated to be above zero with a mean of 1.548 and the confidence interval [1.105 1.919]. This indicates that those IVF-patients where the ICSI fertilization method is used are more likely to have a Fertilization rate above 50 %.

The final model exhibits negative correlation between the target variable, fertilization group, and both the number oocytes and BMI respectively. This indicates that female patients with a high BMI are less likely to have a Fertilization Rate above 50 %. However, this also indicates that the more Ooctyes that were retrieved the lower is the chance to achieve a Fertilization rate above 50 %.

The distribution of the performance measures of the final model is visualized as histograms in figure 4. Note that in all histograms, the performance measures look close to normal distributed, centered around their respective means.

in figure 5, fertilization group is plotted against the LDA-score from the final model. The separation is far from perfect, but most couples in fertilization group 1 ((fertilization rate $\geq 0.5$) have positive LDA-scores and most couples in fertilization group 0 (fertilization rate $< 0.5$) have negative LDA-scores. Since LDA-scores higher than zero are classified as fertilization group 1 and scores lower than zero are classified as fertilization group 0, this means that most samples are correctly classified.



(a) Accuracy

(b) F1-score

(c) Sensitivity

(d) Specificity

Figure 4: Histogram of performance measures for the final model predicting the fertilization group, pairwise subtracted by that of the naive model. The final model only includes the significant parameters fertilization method 2, oocytes, and BMI. The performance measures are retrieved from the bootstrap loop.

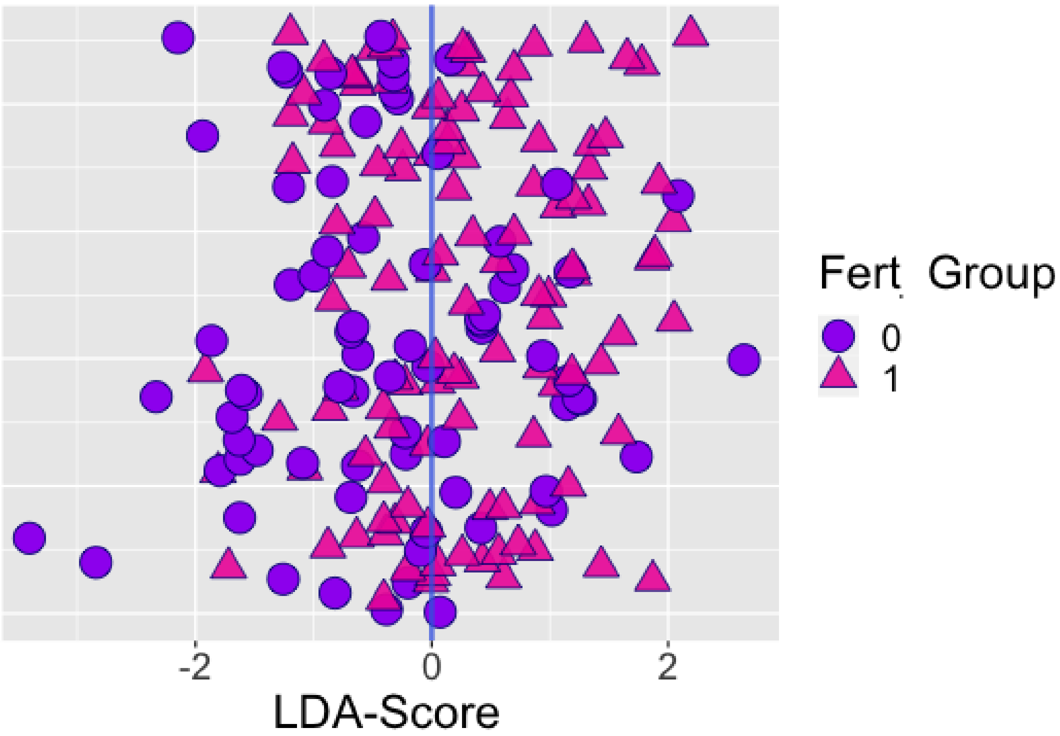
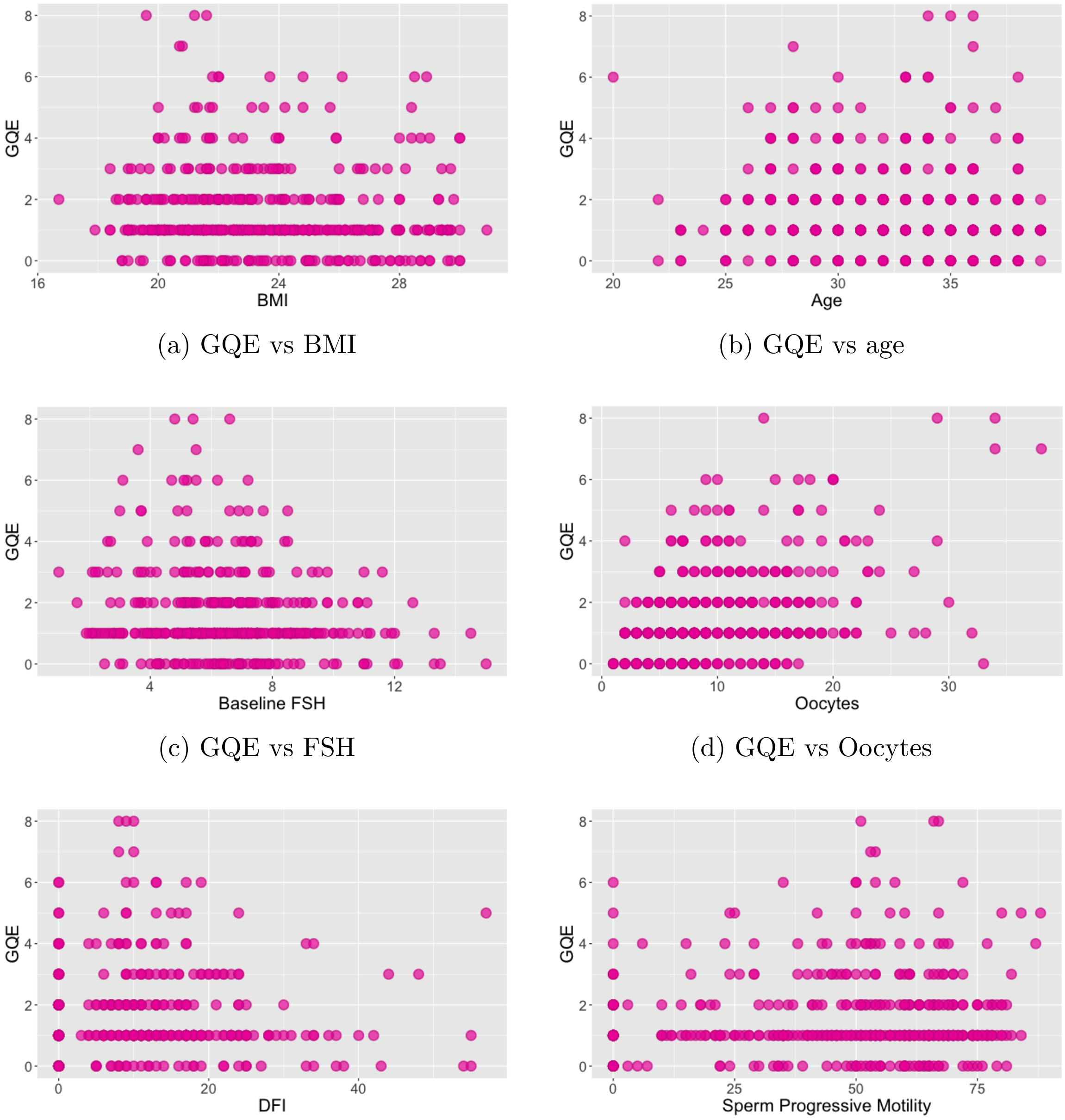


Figure 5: LDA-score for the final model, using fertilization group as target variable and fertilization method 2, BMI and oocytes as model parameters. Positive LDA-scores are classified as 1 (fertilization rate $\geq 0.5$) and negative as 0 (fertilization rate $< 0.5$).

## 10 Prediction of good quality embryos

Moving on to predict the embryo transfer stage, modelling on all couples who had at least one fertilized egg. The number of good quality embryos (GQE) only takes on discrete values between 0-8 in the data set and hence, a linear regression model did not seem appropriate. GQE was also plotted against a few other variables to see if there could be a linear relationship. As can be seen in figure 6, GQE does not seem to have an obvious linear relationship with any of the chosen variables and therefore, fitting linear regression to predict the number of good quality embryos was ruled out.



(a) GQE vs BMI

(b) GQE vs age

(c) GQE vs FSH

(d) GQE vs Oocytes

(e) GQE vs DFI

(f) GQE vs Progressive sperm motility

Figure 6: QGE plotted against multiple variables

In earlier research on the topic, for example in [18], researchers have looked at whether or not the couple gets at least one good quality embryo. Hence, a binary variable called GQE group was created according to the definition in table 15.

| GQE-group: 0 | GQE-group: 1 |
|---|---|
| GQE = 0 | GQE >0 |

Table 15: Categorization of GQE

## 10.1 LDA All Parameters

When a binary variable was formed, it is suitable to use a classification model for the predictions and therefore, an LDA was performed with the GQE group as the target variable. The same method was used as when modeling the fertilization group as both are binary classification problems.

First, all parameters known before the embryo transfer (step 5) were used, see timeline in figure 2, to find the parameters which had the largest absolute normalized weights in the LDA-model. As the weights are normalized, the absolute weights rank the parameters according to their importance for correctly classifying the IVF-patients into their GQE group. The parameters which had an absolute weight larger than 0.05 are exhibited in table 16 in descending order based on their absolute weight.

Table 16 shows that, in descending order, fertilization rate, indication 2, indication 3, fertilization method 2, fertilization method 3, stim Days, sperm volume, the number of fertilized injected eggs, the number of inseminated eggs, BMI, OSI, baseline FSH, and age are the most important parameters when classifying IVF-patients into a GQE group. Only the parameters which had an absolute weight larger than 0.05 are displayed in table 16 as those were the parameters used when creating reduced models. To see the full ranking of all the parameters and their exact absolute weights, see table 38 in the appendix.

| Order | Parameter |
|---|---|
| 1 | Fertilization rate |
| 2 | Indication 2 |
| 3 | Indication 3 |
| 4 | Fertilization method 2 |
| 5 | Fertilization method 3 |
| 6 | Stim days |
| 7 | Sperm Volume |
| 8 | Fertilized injected eggs |
| 9 | Fertilized inseminated eggs |
| 10 | BMI |
| 11 | OSI |
| 12 | Baseline FSH |
| 13 | Age |

Table 16: The parameters of the LDA model predicting the GQE group. The parameters are exhibited in descending order based on their absolute weight. Only parameters with an absolute weight $\geq 0.05$ are included since these were the ones included in the selection of a reduced model. To see the order of all parameters and their absolute weights, see table 38 in section 17.5 in the appendix.

## 10.2 LDA Reduced

Next, the same procedure as in section 9.2 was followed. Starting out, an LDA model which included only the two most important parameters, fertilization rate and indication, was created and then parameters with an absolute weight above 0.05 were systematically added one by one according the order in table 16. The parameter was only kept if it increased the performance according to the criteria explained in section 5.3.

When all parameters in table 16 had been added to the model they were then excluded one by one in the same order that they were added, starting with the most important parameter. If the exclusion of the parameter had a negative impact on the performance of the model it was included again. To see each step of the process where parameters were first added and excluded, see table 39 in the appendix. The table displays the accuracy, F1-score, sensitivity, and specificity of all the models which were tested on the way to find the best performing model.

The performance of best model found after including and excluding variables is exhibited in table 17. The table also includes a naive predictor for reference as the model only adds value if it can significantly outperform the naive predictor. The best model found contained five parameters: fertilization rate, indication, sperm volume, BMI, and OSI.

| Parameters | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|
| Naive predictor | 0.50 [0.40 0.59] | 0.64 [0.55 0.73] | 0.50 [0.39 0.60] | 0.50 [0.10 1.00] |
| Fert. Rate, Indication, Sperm Volume, BMI, OSI | 0.69 [0.59 0.78] | 0.80 [0.72 0.88] | 0.70 [0.58 0.81] | 0.61 [0.20 1.00] |

Table 17: Accuracy, F1-score, sensitivity, and specificity for the naive predictor and the optimal reduced LDA model with the GQE group as target variable. Each performance measure is presented with its mean and its 95% confidence interval.

Coefficients were estimated for all parameters used in the best performing model; the estimates and their respective 95% can be found in table 18. Looking at the coefficient estimates for those parameters, only Fertilization rate and BMI have significant coefficients whose confidence intervals do not cover zero, [1.923 4.253] and [-0.289 -0.059], respectively. Hence, a final model was made using only those two parameters. The performance of this model is exhibited in table 19 and the coefficients of the same model are found in table 20.

| Parameter | Estimate | Conf. int. (95%) |
|---|---|---|
| Fertilization rate | 3.266 | [1.923 4.253] |
| Indication 2 | -0.413 | [-1.698 2.211] |
| Indication 3 | 0.527 | [-0.438 1.563] |
| Indication 4 | -0.059 | [-0.978 1.915] |
| Sperm Volume | 0.168 | [-0.171 0.471] |
| BMI | -0.182 | [-0.289 -0.059] |
| OSI | 0.025 | [-0.051 0.140] |

Table 18: Parameter estimates for the LDA model that best predicted the GQE group, which performance was exhibited in table 17. The estimates are presented as well as their 95% confidence interval.

Table 19 shows that the final LDA model classified with an average accuracy of 0.64 and that the accuracy was significantly higher than that of the naive "flip-a-coin" model, since the confidence interval of the paired difference does not cover zero

([0.02 $\infty$]). The LDA model also outperformed the naive when it comes to F1, with an average score of 0.76. The confidence interval of the F1-scores' paired difference, [0.01 $\infty$], does not cover zero which shows the significance. The final model had a significantly higher sensitivity than did the naive model, shown by the confidence of the paired difference which does not cover zero ([0.01, $\infty$]). The specificity's confidence interval of the paired difference, [-0.25 $\infty$], does cover zero though. Hence, the final model does not have a significantly higher specificity than does the naive model.

Note the high uncertainty of the specificity; The confidence interval of the naive model reaches all the way up to 1.00. It is impossible to get a coefficient estimate that lies significantly above the interval. This problem comes from the observations of couples not getting at least one good quality embryo being too few. In conclusion, table 19 shows that the final model significantly outperforms the naive predictor in all performance measures except for specificity.

| Model/<br>Comparison | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|
| Naive predictor | 0.50 [0.40 0.59] | 0.64 [0.55 0.73] | 0.50 [0.39 0.60] | 0.50 [0.10 1.00] |
| Fert. rate, BMI | 0.64 [0.55 0.73] | 0.76 [0.68 0.83] | 0.64 [0.52 0.75] | 0.67 [0.33 1.00] |
| Paired difference | 0.14 [0.02 $\infty$] | 0.12 [0.01 $\infty$] | 0.14 [0.01 $\infty$] | 0.18 [-0.25 $\infty$] |

Table 19: Accuracy, F1-score, sensitivity, and specificity for the naive predictor and the final LDA model when only including significant parameters, still with GQE group as target variable. Each performance measure is presented with its mean and its 95% confidence interval. The paired difference between the final and naive model is also included with its lower limited 95% confidence interval.

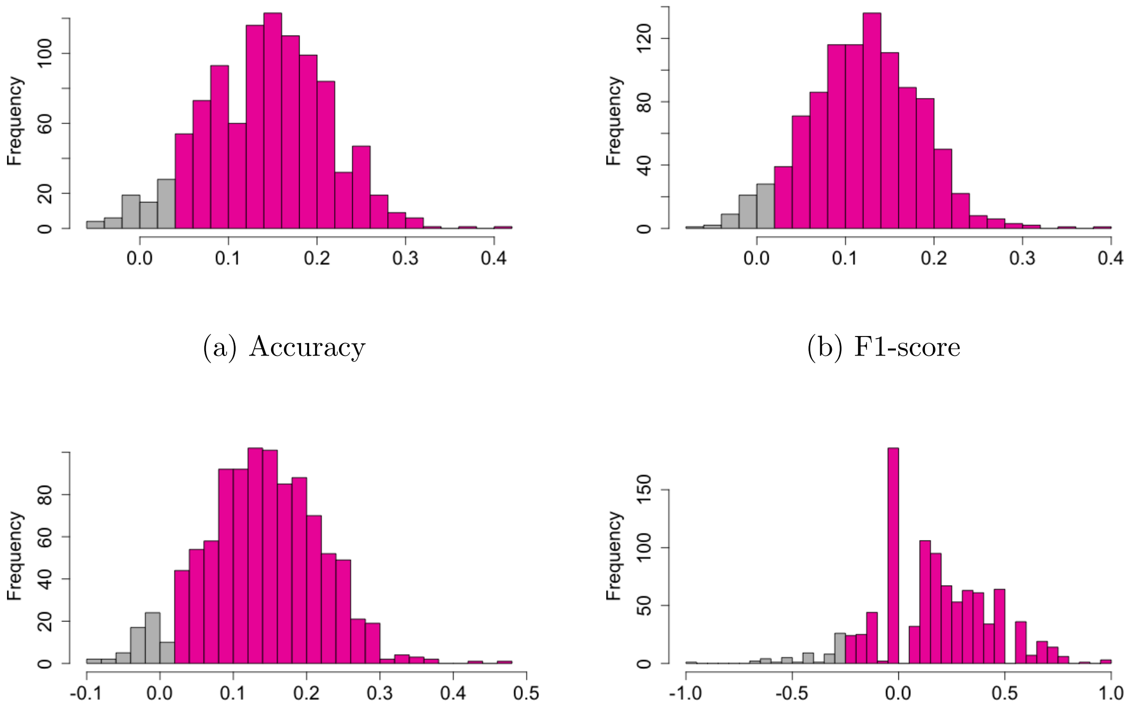
The parameter estimates of the final model, which are shown in table 20, indicate that there is a positive correlation between fertilization rate and obtaining at least one good quality embryo and a negative correlation between BMI and obtaining at least one good quality embryo.

| Parameter | Estimate | Conf. int. (95%) |
|---|---|---|
| Fertilization Rate | 3.671 | [2.252 4.567] |
| BMI | -0.204 | [-0.312 -0.080] |

Table 20: Parameter estimates for the LDA model predicting fertilization group using only significant parameters, which performance was exhibited in table 19. The estimates are presented as well as their 95% confidence interval.

The performance of the final model is visualized as histograms in figure 7. Accuracy, F1, and sensitivity look fairly normally distributed around their means. Specificity on the other hand looks messier, which is because there are so few observations of GQE group 0 (no good quality embryos).

in figure 8, the GQE group is plotted against the LDA-score from the final model. Perfect separation is not achieved, but most couples in GQE group 1 (at least one good quality embryo) have positive LDA-scores and most couples in GQE group 0 (no good quality embryos) have negative LDA-scores. Since LDA-scores higher than zero are classified as GQE group 1 and scores lower than zero are classified as GQE group 0, this means that most samples are correctly classified. Note how few observations there are of GQE group 0, represented by the circles in the figure.



(a) Accuracy

(b) F1-score

(c) Sensitivity

(d) Specificity

Figure 7: Histogram of performance measures for the final model, pairwise subtracted by that of the naive model, predicting the GQE group. The final model only includes the significant parameters BMI and fertilization rate. The performance measures are retrieved from the bootstrap loop.

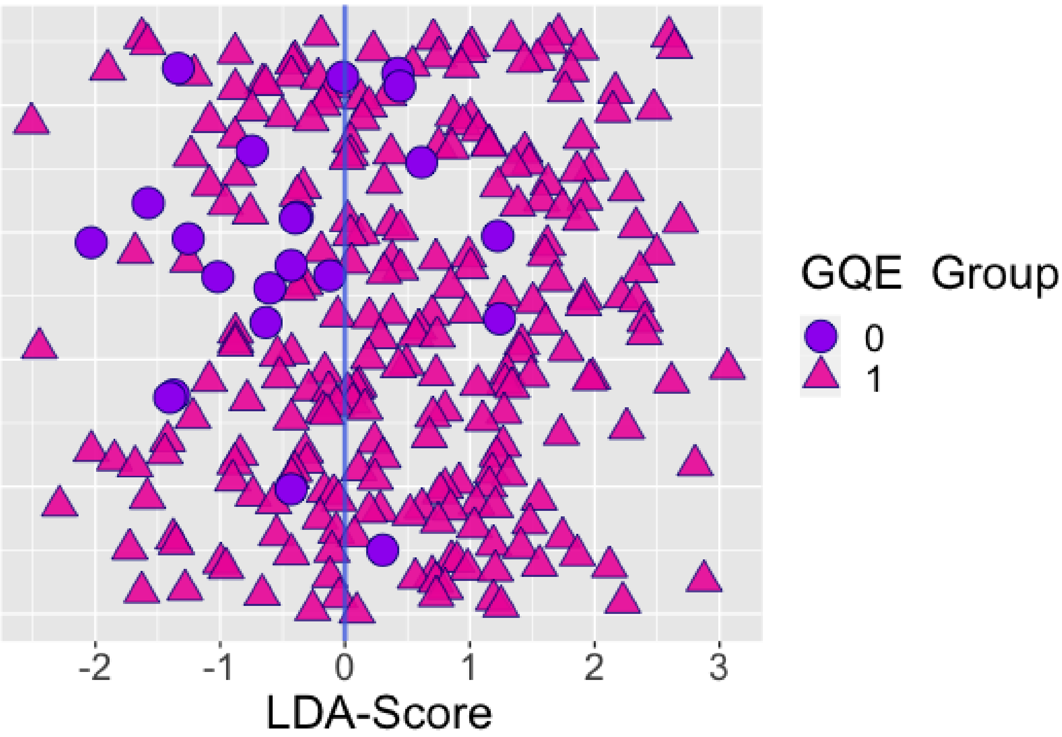


Figure 8: LDA-score for the final model, using fertilization group as target variable and with fertilization rate and BMI as model parameters. Positive LDA-scores are classified as 1 (at least one good quality embryo) and negative as 0 (no good quality embryos).

# 11 Prediction of pregnancy

## 11.1 LDA All Parameters

When predicting the next step of the treatment, pregnancy (step 6), all couples who received at least one good quality embryo were included in the data set and all pregnancies were treated equally, meaning that the couple was classified as pregnant both if they used a fresh or a frozen embryo. Data wise, this means that the Pregnant Fresh and the Pregnant FER parameters were added to create the variable Pregnancy, which was then used as the target variable. As Pregnancy is a variable with a binary outcome the LDA model was chosen.

All parameters which were known before the pregnancy (step 6), see timeline in figure 2, were included in a first LDA model. Table 21 shows the order of importance for the parameters with an absolute weight above 0.05. To see the full ranking of all the parameters and their exact absolute weights, see table 40 in the appendix.

| Order | Parameter |
|---|---|
| 1 | Fertilization method 2 |
| 2 | Fertilization method 3 |
| 3 | Indication 3 |
| 4 | GQE |
| 5 | Fertilized injected eggs |
| 6 | Indication 2 |
| 7 | ET Day |
| 8 | Fertilized inseminated eggs |
| 9 | Baseline FSH |
| 10 | Sperm Volume |
| 11 | Cycle Length |
| 12 | BMI |
| 13 | Indication 4 |

Table 21: The parameters of the LDA model predicting pregnancy. The parameters are exhibited in descending order based on their absolute weight. Only parameters with an absolute weight $\geq 0.05$ are included since these were the ones included in the selection of a reduced model. To see the order of all parameters and their absolute weights, see table 40 in section 17.6 in the appendix.

## 11.2 LDA Reduced

After creating the full model and getting the order of parameter weights exhibited in table 21, a reduced model using fewer parameters and performing better than the initial one was sought. The same procedure as in sections 9.2 and 10 was performed again, first adding and then removing parameters systematically. To see each step of the process where parameters were first added one by one in order of importance and then excluded in the same order, see table 41 in the appendix. The table displays the accuracy, F1-score, sensitivity, and specificity of all the models which were tested on the way to find the best performing model.

The best reduced model contained four parameters: fertilization method, indication, fertilized injected eggs, and BMI, and the model's performance is exhibited in table 22. The table also includes a naive predictor for reference. The coefficients of the best model are exhibited in table 23. Both fertilization methods, indication 3, the number of fertilized injected eggs, and BMI had significant coefficients whose confidence intervals do not cover zero while Indication 2 and 4 have insignificant coefficients whose confidence intervals cover zero.

| Model | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|
| Naive predictor | 0.50 [0.39 0.61] | 0.518 [0.39 0.64] | 0.50 [0.35 0.63] | 0.51 [0.35 0.67] |
| Fertilization method, Indication, BMI, Fertilized injected eggs | 0.56 [0.49 0.71] | 0.64 [0.54 0.72] | 0.72 [0.54 0.90] | 0.38 [0.19 0.60] |

Table 22: Accuracy, F1-score, sensitivity, and specificity for the naive predictor and the optimal reduced LDA model with pregnancy as target variable. Each performance measure is presented with its mean and its 95% confidence interval.

| Parameter | Estimate | Conf. int. (95%) |
|---|---|---|
| Fertilization method 2 | -2.580 | [-3.394 -1.481] |
| Fertilization method 3 | -1.998 | [-3.323 -0.488] |
| Indication 2 | -0.292 | [-2.096 1.599] |
| Indication 3 | 1.684 | [ 0.871 2.444] |
| Indication 4 | 0.744 | [-0.468 1.970] |
| Fertilized injected eggs | 0.263 | [ 0.068 0.427] |
| BMI | -0.167 | [-0.266 -0.049] |

Table 23: Parameter estimates for the LDA model that best predicted pregnancy, which performance was exhibited in table 22. The estimates are presented as well as their 95% confidence interval.

As the parameter estimates in table 23 show that indication 2 and 4 are , a dummy variable was created for indication 3 (male factor), to exclude the insignificant indication 2 and 4 from the model. It was created as defined by table 24.

| Indication 3 dummy: 0 | Indication 3 dummy: 1 |
|---|---|
| Indication $\neq 3$ | Indication $= 3$ |

Table 24: Defining a dummy variable for fertilization method 2

The coefficients of the final model, only including the significant parameters fertilization method 2 and 3, indication 3, the number of fertilized injected eggs, and BMI, is exhibited in table 26. The performance of the same model is shown in table 25, where the naive predictor is included for reference as well as the test statistic paired difference.

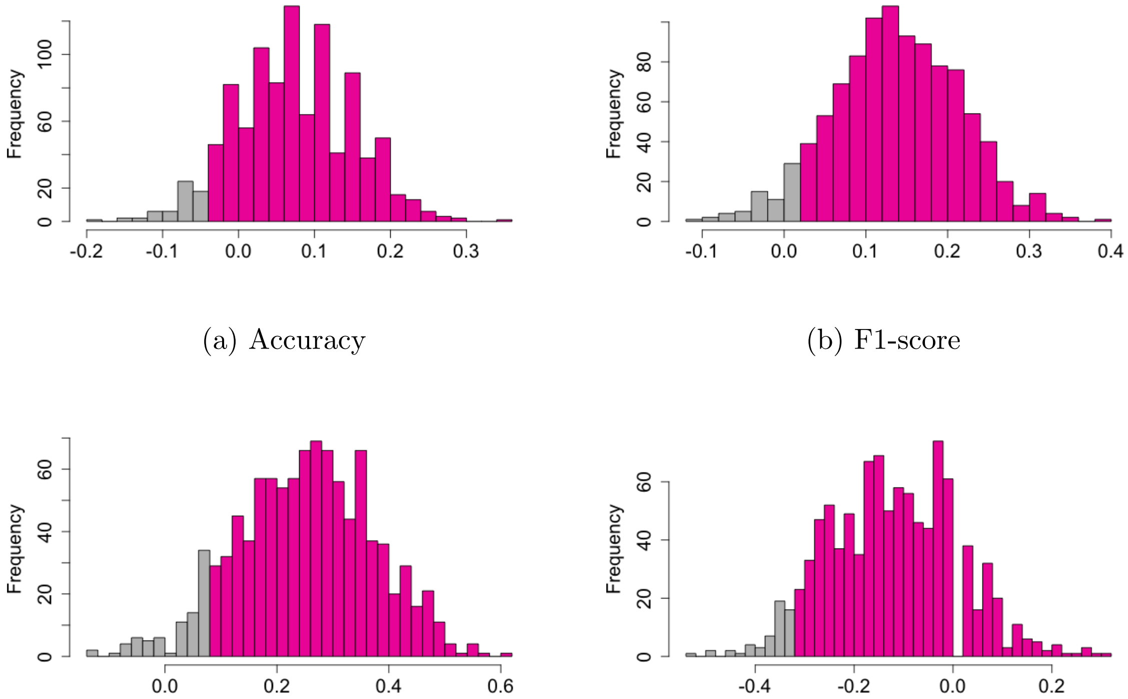
| Model/ Comparison | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|
| Naive predictor | 0.50 [0.39 0.61] | 0.52 [0.39 0.64] | 0.50 [0.35 0.63] | 0.51 [0.35 0.67] |
| Fert. method Fert. inj. eggs, Indication 3, BMI | 0.57 [0.48 0.68] | 0.65 [0.56 0.74] | 0.75 [0.55 0.91] | 0.38 [0.19 0.59] |
| Paired difference | 0.07 [-0.05 $\infty$] | 0.14 [0.01 $\infty$] | 0.25 [0.06 $\infty$] | -0.12 [-0.32 $\infty$] |

Table 25: Accuracy, F1-score, sensitivity, and specificity for the naive predictor and the final LDA model when only including significant parameters, still with pregnancy as target variable. Each performance measure is presented with its mean and its 95% confidence interval. The paired difference between the final and naive model is also included with its lower limited 95% confidence interval.

| Parameter | Estimate | Conf. int. (95%) |
|---|---|---|
| Fertilization method 2 | -2.671 | [-3.470 -1.615] |
| Fertilization method 3 | -2.132 | [-3.477 -0.569] |
| Indication 3 | 1.701 | [0.870 2.410] |
| Fertilized injected eggs | 0.264 | [0.059 0.431] |
| BMI | -0.172 | [-0.273 -0.051] |

Table 26: Parameter estimates for the LDA model predicting fertilization group using only significant parameters, which performance was exhibited in table 25. The estimates are presented as well as their 95% confidence interval.

The final model outperforms the naive ”flip-a-coin” model when it comes to F1-score and sensitivity, since the confidence intervals of their paired differences do not cover zero ([0.01 $\infty$] and [0.06 $\infty$], respectively). However, the final model does not have a significantly higher accuracy nor specificity than does the naive model. This is seen since the confidence intervals of the paired differences, [-0.05 $\infty$] and [-0.32 $\infty$], cover zero. Note, that this is a balanced data set. Hence, the uncertainty of the specificity is comparable to that of the sensitivity in this case.



(a) Accuracy

(b) F1-score

(c) Sensitivity

(d) Specificity

Figure 9: Histogram of performance measures for the final model pairwise subtracted by that of the naive model predicting pregnancy. The final model only uses the significant parameters fertilization method, fertilized injected eggs, Indication 3 and BMI. The performance measures are retrieved from the bootstrap loop.

# 12 Prediction of live birth

## 12.1 LDA All Parameters

Now, to the final step of the IVF treatment, step 7 in the timeline found in figure 2. In this step, only the couples who got pregnant were included in the data set. The target variable is Live Birth (LB), which is whether or not a living child was born. As with QGE Group and Total Pregnancies, this variable is a binary variable and therefore the LDA classification method was chosen to model Live Birth.

The LDA procedure was repeated once again starting with a model containing all parameters known before step 7, and then building a reduced model using the weights of the large model and the performance of the reduced to make decisions whether or not to include the parameter in the model. The parameters with absolute weights above 0.05 a are exhibited in table 27 in descending order according to their absolute weight. To see the full ranking of all the parameters and their exact absolute weights, see table 42 in the appendix.

| Order | Parameter |
|---|---|
| 1 | Stim Days |
| 2 | Fertilized inseminated eggs |
| 3 | BMI |
| 4 | ET Day |
| 5 | GQE |
| 6 | OSI |
| 7 | Fertilized injected eggs |
| 8 | Cycle Length |
| 9 | Sperm Volume |

Table 27: The parameters of the LDA model predicting live birth. The parameters are exhibited in descending order based on their absolute weight. Only parameters with an absolute weight $\geq 0.05$ are included since these were the ones included in the selection of a reduced model. To see the order of all parameters and their absolute weights, see table 42 in section 17.7 in the appendix.

## 12.2 LDA Reduced

| Model/<br>Comparison | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|
| Naive predictor | 0.50 [0.35 0.63] | 0.61 [0.46 0.75] | 0.50 [0.34 0.66] | 0.50 [0.14 0.80] |
| BMI, GQE | 0.56 [0.43 0.67] | 0.65 [0.49 0.78] | 0.53 [0.34 0.71] | 0.68 [0.30 1.00] |
| Paired difference | 0.06 [-0.11 $\infty$] | 0.05 [-0.13 $\infty$] | 0.03 [-0.17 $\infty$] | 0.17 [-0.25 $\infty$] |

Table 28: Accuracy, F1-score, sensitivity, and specificity for the naive predictor and the final LDA model with live birth as target variable. Each performance measure is presented with its mean and its 95% confidence interval. The paired difference between the final and naive model is also included with its lower limited 95% confidence interval.

After including and excluding parameters, the best LDA model found for predicting live birth included only two parameters: GQE and BMI, see table 28. The table also contains a naive predictor for reference. To see each step of the process where parameters were first added systematically in order of importance and then later excluded in the same order see table 43 in the appendix. Looking closer at the estimated coefficients of the parameters seen in table 29, only the coefficient for GQE is significant as the confidence interval does not cover zero, [0.070 0.853], while that for BMI does, [-0.028 0.378].

When looking at the performance of the best model, which is exhibited in table 28, it does not outperform the naive model in any of the performance categories. This is shown by the confidence intervals of the paired differences, which all cover zero.

| Parameter | Estimates | Conf. int. (95%) |
|---|---|---|
| BMI | 0.222 | [-0.028 0.378] |
| GQE | 0.481 | [0.070 0.853] |

Table 29: Parameter estimates for the LDA model that best predicted live birth, which performance was exhibited in table 28. The estimates are presented as well as their 95% confidence interval.

# 13 Testing the models on the test data set

In this section, all the final models of each step of the treatment was tested on the test data set. The full modelling set was used to estimate the coefficients of the models, still excluding samples based on exclusion criteria and missing samples. The naive models, an intercept for the linear regression and a "flip-a-coin" model for the LDA, were used on the exact same test data as the final models, excluding the same samples. The performance between the final and the naive models were then compared.

The linear regression model predicting the number of oocytes (step 3 in the timeline in figure 2) predicted the outcome of the test data set with a MSE of 4.56. This was smaller than that of the naive model which was 5.04. The MSE of the linear models are exhibited in table 30.

in table 31, the performance of all LDA models is exhibited. The two LDA models which significantly outperformed the naive models when using the random training and validation sets, predicting fertilization group (step 4) and GQE group (step 5), both performed better than the naive model on all performance measures except specificity. The final model predicting fertilization group had an accuracy and F1-score of 0.63 and 0.72, respectively, while those numbers were 0.45 and 0.51 for the naive model. The sensitivity of the final model for fertilization was 0.77 while that of the naive was 0.46, and the specificity was 0.42 and 0.38, respectively. The LDA model predicting GQE group performed with an accuracy and an F1-score of 0.62 and 0.75, respectively, while the naive only performed with an accurracy of 0.49 and an F1-score of 0.64. The sensitivity of the model predicting GQE was 0.63 while that of the naive model was 0.49. Finally, the specificity of both the final GQE model and the naive model was 0.40.

The best LDA model predicting pregnancy (step 6) did not perform significantly better than the naive model when using the random training and validation sets. It did not convincingly outperform the naive model when predicting the outcome of test set either. The LDA model predicted with an accuracy of 0.48 while the naive model had an accuracy of 0.49. The LDA model had a slightly higher F1-score, 0.57, than did the naive model, 0.53. The sensitivity was higher for the LDA model, 0.61 and 0.52 for the two models respectively, but the specificity was lower, 0.31 and 0.45 respectively.

The last LDA model, predicting live birth (step 7), did not perform significantly better than the naive model either when using the random training and validation sets. However, it outperformed the naive predictor on all performance measures when predicting the outcome of the test set. The LDA model exhibited an accuracy and an F1-score of 0.60 and 0.69, respectively, while the naive model only performed with accuracy and F1 of 0.45 and 0.56. The sensitivity and specificity of the LDA model were 0.63 and 0.52, respectively, while those of the naive were 0.48 and 0.38.

| Model | MAE |
|---|---|
| Oocytes naive | 5.043 |
| Oocytes final | 4.568 |

Table 30: Mean absolute error of the naive and final model predicting the number of retrieved oocytes.

| Model | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|
| Fert. group naive | 0.447 | 0.509 | 0.462 | 0.421 |
| Fert. group final | 0.627 | 0.720 | 0.774 | 0.386 |
| GQE group naive | 0.489 | 0.644 | 0.496 | 0.400 |
| GQE group final | 0.617 | 0.755 | 0.634 | 0.400 |
| Pregnancy naive | 0.488 | 0.529 | 0.514 | 0.455 |
| Pregnancy final | 0.480 | 0.570 | 0.614 | 0.309 |
| LB naive | 0.452 | 0.556 | 0.481 | 0.381 |
| LB final | 0.603 | 0.695 | 0.635 | 0.524 |

Table 31: Accuracy, F1-score, sensitivity, and specificity of the naive and final LDA models predicting fertilization group (Fert. group), GQE group, pregnancy, and live birth (LB).

# 14 Discussion

When revisiting the previous research done on this data set [18], examining the interaction between DFI and baseline AMH, the significance of DFI was confirmed. The significance of the interaction which was found in [18] was however not significant in this study. This is likely because the imputation of the missing values of DFI in the previous study resulted in some correlations which did not reflect the true relationships of the parameters. Since no significance was found when the true DFI values were used, there is reason to reevaluate the conclusion that the interaction between baseline AMH and DFI significantly affects whether or not the couple obtains at least one good quality embryo.

One reason why the results differed between this study and [18] could be how the imputation was computed. 76 out of the total of 204 couples which underwent a standard IVF treatment had missing DFI values in [18], which leaves only 128 data points to fit the imputation model. The hypothesis that the interaction affects the number of good quality embryos should be tested on a larger data set to achieve lower uncertainty in both potential imputations and more importantly the parameter estimates of the AMH and DFI interaction term.

When predicting the number of retrieved oocytes, this study exhibited a significant positive impact of age on the number of retrieved oocytes, i.e., more oocytes are retrieved from older women than from younger. This is the opposite from what previous studies have shown, for example [7] and [14] found that the number of retrieved oocytes significantly decreased with age.

To investigate if there were clear outliers which could affect the results in such a way, oocytes were plotted against age in figure 10. No obvious outliers were found but a sensitivity test was made, fitting a new regression when excluding all couples with more oocytes than 25 (N=10). Even when excluding those with the most oocytes, the coefficient was significantly greater than zero with a confidence interval of [0.063 0.315]. Hence, removing these potential outliers did not remove the significance which goes against previous studies.

in figure 10, samples above the line at 25 oocytes represent the couples which were excluded in the sensitivity test explained above. The continuous regression line represents the regression which was fitted to the data set including all numbers of oocytes, while the dashed line shows the regression fitted to the reduced data set which excludes the couples with more than 25 oocytes. As expected, the dashed line lies below the continuous one. However, both lines are increasing, exhibiting the positive correlation between age and the number of oocytes which goes against previous research.

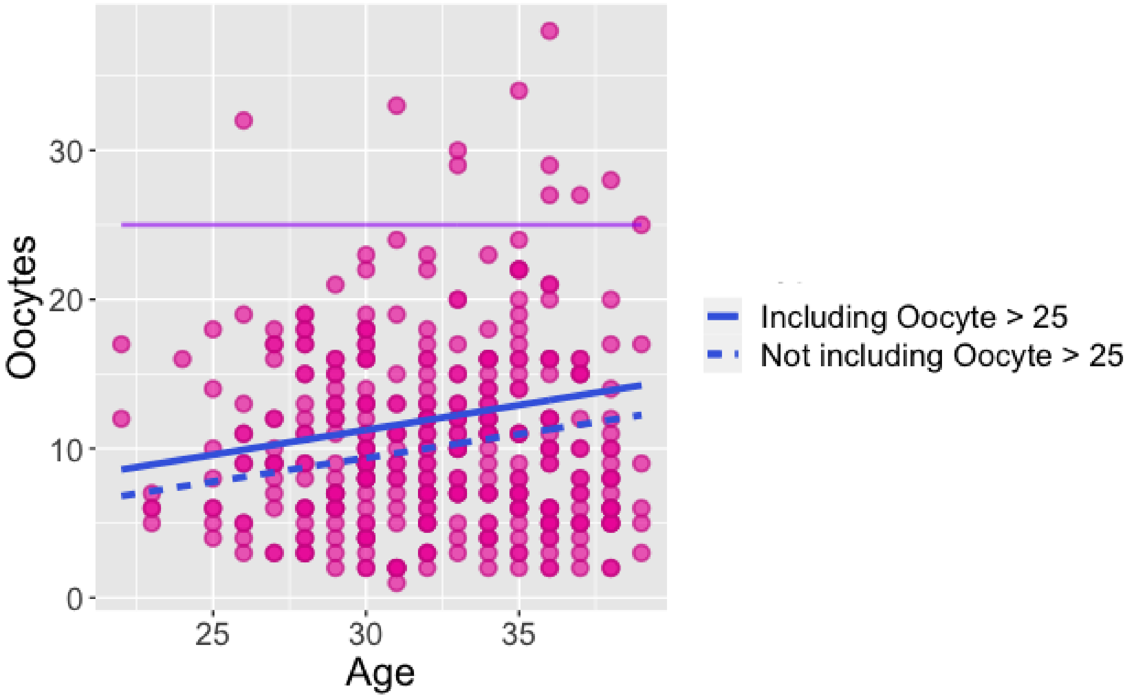


Figure 10: Plot showing correlation between number of retrieved oocytes and age. The samples above the line at 25 oocytes represent the couples which were excluded in the sensitivity test. The continuous regression line represents the regression which was fitted to the data set including all numbers of retrieved oocytes, while the dashed line shows the regression fitted to the reduced data set which excludes the couples with more than 25 retrieved oocytes.

In the final model predicting oocytes, baseline AMH and FSH levels were also included. AMH was significantly positive and FSH was significantly negative, meaning that increased levels of baseline AMH increases the number of oocytes while higher levels of baseline FSH have the opposite effect. This is well aligned with previous studies, such as [8] showing a positive correlation between AMH and collected oocytes and [1] showing a negative correlation between FSH and collected oocytes.

When predicting fertilization group, GQE, and pregnancy, increasing BMI had a significant negative impact on the outcome. Study [14] showed a significant negative correlation between BMI and the probability of pregnancy, which is well aligned with the results of this study. Those studies that found significance of BMI when it comes to fertilization rate and embryo quality, for example [15] [12] [4] [9], also found negative correlations between BMI and the target variable. The results of this study strengthen the earlier findings of a negative impact of BMI on the outcome of the treatment.

The number of collected oocytes was significant when predicting fertilization group, and fertilization rate was significant when predicting good quality embryos. Hence, these parameters can be used as early indicators of the ongoing treatment, potentially opening up the possibility of adapting the treatment based on the predictions. No previous studies finding significance between number of collected oocytes and fertilization group were found. In [13], the correlation was sought but no significance was found. No previous studies were found which investigated the correlation between fertilization rate and good quality embryos, respectively. Hence, these results add new information about prediction of IVF treatments which could be valuable in future research and potentially in clinical practice.

The number of collected oocytes exhibited a negative correlation with fertilization rate, meaning that more oocytes leads to lower fertilization rate which might seem counter-intuitive. When searching for previous research no research was found which specifically covers the topic of the correlation between fertilization rate and the number of retrieved oocytes. More research should be done to confirm or refute the found correlation.

The final model predicting live birth only had one significant parameter: the number of good quality embryos. BMI was also in the final model but was not significant. In previous studies, for example [5], it was however shown to be significant. The data set in this current study was small (N=172) after excluding all couples which did not get pregnant, which resulted in high uncertainty of the parameter estimates and hence low significance. When testing the final model on the test data set, however, it outperformed the naive model on all performance measures. Hence, it could be valuable to revisit these parameters when modelling a larger data set. No previous studies were found that investigated the correlation between good quality embryo count and pregnancy rate, making it interesting to examine the significance of GQE which was found in this study, on a larger data set.

# 15 Conclusion

In this final section, a concise review of the main results of the study as well as how it relates to previous research is presented.

When revisiting the interaction between baseline AMH (anti-mullerian hormone) and DFI (DNA fragmentation index), the significance on the GQE group (recieveing at least one good quality embryo) which was found in [18] could not be confirmed. The loss of significance most likely comes from the addition of true DFI values; imputation was used in [18] for 76 missing values of DFI, while all true values were known in this study. Further research needs to be done on larger data sets to investigate whether or not the interaction between baseline AMH and DFI affects the embryo quality.

Baseline AMH, baseline FSH, and age, were shown to affect the number of retrieved oocytes. A positive correlation was found with baseline AMH and a negative with baseline FSH. Both theses results are in line with previous studies on the topic. Age however exhibited a positive correlation with the number of retrieved oocytes, which goes against previous research and the intuition that older women have fewer eggs. This is an important result, especially since age is one of the criteria which determines whether a couple is eligible for IVF treatments. Hence, it should be investigated further on other data sets.

It was found that high BMI had a significant negative impact on fertilization rate. High BMI was also shown to have a significant negative impact on the chance of receiving at least one qood quality embryo, among the patients which had at least one fertilized egg. This is both in line with previous research and further strengthen the idea that higher BMI has a negative impact on IVF outcome.

Finally, it was shown that there is a predictive value in the number of collected oocytes when it comes to predicting fertilization group, and that fertilization rate can significantly help predicting the GQE group (receiving at least one good quality embryo). No previous research was found which exhibits these results. This new information could add value in the future development of predictive models for IVF treatments.

# 16 References

[1] H. Abdalla, M.Y. Thum. *An elevated basal FSH reflects a quantitative rather than qualitative decline of the ovarian reserve.* Human Reproduction, Volume 19, Issue 4, April 2004, Pages 893–898.

[2] S.S.E. Alson, L.J. Bungum, A. Giwercman, E. Henic, *Anti-müllerian hormone levels are associated with live birth rates in ART, but the predictive ability of anti-müllerian hormone is modest* European Journal of Obstetrics & Gynecology and Reproductive Biology, Volume 225, 2018, Pages 199-204.

[3] K. Bengtsson. "IVF, Provrörsbefruktning", 1177 Vårdguiden. 2020-01-31.

[4] D.T. Carrell, K.P. Jones, C.M. Peterson, V. Aoki, B.R. Emery, B.R. Campbell. *Body mass index is inversely related to intra-follicular HCG concentrations, embryo qualityand IVF outcome.* Reproductive BioMedicine Online webpaper, Volume 3, No 2, 2001, Pages 109–111.

[5] P. Fedorcsák, P.O. Dale, R. Storeng, G. Ertzeid, S. Bjercke, N. Oldereid, A.K. Omland, T. Åbyholm, T. Tanbo. *Impact of overweight and underweight on assisted reproduction treatment.* Human Reproduction, Volume 19, Issue 11, November 2004, Pages 2523–2528, https://doi.org/10.1093/humrep/deh485.

[6] B. Friberg & M. Kitlinski. *Regional riktlinje för assisterad befruktning*, Region Skåne. 2021-03-04.

[7] L. Janny & Y. J.R. Menezo. *Maternal Age Effect on Early Human Embryonic Development and Blastocyst Formation*, Molecular Reproduction and Development, Issue 45, 1996.

[8] L. Kotanidis, K. Nikolettos, S. Petousis, B. Asimakopoulos, E. Chatzimitrou, G. Kolios & N. Nikolettos. *The use of serum anti-Mullerian hormone (AMH) levels and antral follicle count (AFC) to predict the number of oocytes collected and availability of embryos for cryopreservation in IVF.* Journal of Endocrinological Investigation, Volume 39, 2016, p. 1459–1464.

[9] M. Metwally, R. Cutting, A. Tipton, J. Skull, W.L. Ledger, T.C. Li. *Effect of increased body mass index on oocyte and embryo quality in IVF patients.* Reproductive BioMedicine Online, Volume 15, Issue 5, 2007, Pages 532-538.

[10] Q-IVF. *Fertilitetsbehandlingar i Sverige - Årsrapport 2020.* 2020.

[11] L. Sabatini, A. Zosmer, E.M. Hennessy, A. Tozer, T. Al-Shawaf, *Relevance of basal serum FSH to IVF outcome varies with patient age* Reproductive BioMedicine Online, Volume 17, Issue 1, 2008, Pages 10-19.

[12] O. Salha, T. Dada  V. Sharma. *Influence of body mass index and self-administration of hCG on the outcome of IVF cycles: A prospective cohort study.* Human Fertility, Volume 4, Issue 1, 2001, pages 37-42.

[13] F.I. Sharara, H.D. McClamrock. *High estradiol levels and high oocyte yield are not detrimental to in vitro fertilization outcome.* Fertility and Sterility, Volume 72, Issue 3, September 1999, Pages 401-405.

[14] M.L. Sneed, M.L. Uhler, H.E. Grotjan, J.J. Rapisarda, K.J. Lederer, A.N. Beltsos. *Body mass index: impact on IVF success appears age-related.* Human Reproduction, Volume 23, Issue 8, August 2008, Pages 1835–1839.

[15] E.C.A.M. van Swieten, L. van der Leeuw-Harmsen, E.A. Badings, P.J.Q. van der Linden. *Obesity and Clomiphene Challenge Test as Predictors of Outcome of in vitro Fertilization and Intracytoplasmic Sperm Injection.* Gynecol Obstet Invest, Volume 59, 2005, Pages 220-224.

[16] Z. Wang, H. Groen, K.C. Van Zomeren, A.E.P. Cantineau, A. Van Oers, A.P.A. Van Montfoort, W.K.H. Kuchenbecker, M.J. Pelinck, F.J.M. Broekmans, N.F. Klijn, E.M. Kaaijk, B.W.J. Mol, A. Hoek, J. Van Echten-Arends. *Lifestyle intervention prior to IVF does not improve embryo utilization rate and cumulative live birth rate in women with obesity: a nested cohort study.* Human Reproduction Open, Volume 2021, 2021, Issue 4.

[17] D.L. Zander-Fox, R. Henshaw, H. Hamilton, M. Lane. *Does obesity really matter? The impact of BMI on embryo quality and pregnancy outcomes after IVF in women aged $\leq$ 38 years.* Aust N Z J Obstet Gynaecol, volume 52, issue 3, June 2012, pages 270-276.

[18] P. Zarén, S. Alson, E.Henic, M. Bungum, A. Giwercman *Interaction between serum levels of Anti-Mullerian Hormone and the degree of sperm DNA fragmentation measured by sperm chromatin structure assay can be a predictor for the outcome of standard in vitro fertilization*, PLoS ONE 14(8). 2019-08-08.

# 17 Appendix

## 17.1 Parameter Lexicon

**1. BMI** - Body Mass Index of the Female Patient. BMI is calculated as

$$\text{BMI} = \frac{\text{weight [kg]}}{\text{height [m]}^2} \tag{11}$$

**2. Age** - The age of the female patient in Years.

**3. Follicles** - Follicle count. Follicles are fluid filled sacs located in the ovaries. The female sex cell (egg cell) develop within the follicles and each follicle can hold one egg.

**4. Oocytes** - Number of retrieved oocytes. An oocyte is an immature egg cell located within a follicle.

**5. Fertilization method** - Method chosen to fertilize the mature eggs from the female patient with the sperm from the male patient. In the data set used for this report, three different fertilization methods have been used on the patients. Fertilization method 0 indicates those who went through a Standard IVF, where sperm and egg are placed together for fertilization to occur. Fertilization method 1 indicates those couples were ICSI fertilization was used. During ICSI a single sperm is injected into the egg for fertilization to occur. Fertilization method 2 indicates a microcombonation, which is a combination between ICSI and Standard IVF.

**6. Fertilization rate** - fertilization rate is defined as the number of eggs on which fertilization was succesful divided by the total number of eggs on which fertilization was attempted

**7. ET Day** - Day of embryo transfer. Embryo transfer is when an embryo (a fertilized egg in the early stages of development) is transferred into the female patients uterus.

**8. GQE** - GQE count, where GQE stands for Good Quality Embryo.

**9. GQE per Oocyte** - Ratio of good quality embryos divided by the oocyte count.

**10. OSI** - OSI stands for Ovarian Sensitivity Index and is calculated as

$$\text{OSI} = \frac{\text{Number of Retrieved Oocytes}}{\text{Total gonadotropin Dose}} \times 1000 \tag{12}$$

**11. Pregnant Fresh** - Total number of pregnancies reported among patients using fresh embryos

**12. Pregnant FER** - Total number of pregnancies reported among patients using FER (Frozen Embryo Replacement)

**13. Pregnancy** - Total number of pregnancies reported among patients. Pregnancy is calculated as pregnant fresh + pregnant FER

**14. Spontaneous abortion before Week 6** - Total number of spontaneous abortions reported among patients before week 6

**15. Spontaneous abortion in Week 6-12** - Total number of spontaneous abortions reported among patients between week 6-12

**16. Live Birth Fresh** - Total number of live births reported among patients using fresh embryos

**17. Live Birth FER** - Total number of live births reported among patients using FER (Frozen Embryo Replacement)

**18. Live Birth** - Total number of live births reported among patients. It is calculated as live birth fresh + live birth FER

**19. Stim Days** - Number of days gonadotropin (fertility drug) is given to the female patient

**20. Initial Dose** - Initial gonadotropin dose given to the female patient

**21. Total Dose** - Total gonadotropin dose given to the female patient

**22. Dose per Day** - Dose per day of gonadotropin given to the female patient

**23. Indication** - Indication key stating the reason for infertility. Note that this is an evaluation made by a doctor and is therefor subjective. In this data set, 4 different indications were reported among patients. They are labeled Indication 1, Indication 2, Indication 3, and Indication 4. Indication 1 indicates that the reason for infertility is unexplainable, indication 2 and 3 indicates that the reason for infertility is due to female and male parameters, respectively, and indication 4 indicates that the reason for infertility is due to a mix of male and female conditions.

**24. PCOS** - Parameter indicating whether the female patient has PCOS or not. PCOS stands for Polycystic ovary syndrome, and is a hormonal condition where the ovaries produce an abnormal amount of male hormones. Patient with PCOS can have large number of follicles, however most of these follicles do not release an egg, and therefore irregular periods are common among patients with PCOS. PCOS may cause infertility.

**25. OHSS** - Parameter indicating whether the female patient had Ovarian hyperstimulation syndrome (OHSS) as a reaction of IVF treatment. OHSS is a complication that can occur after IVF as a reaction of the increased hormone levels and it causes the ovaries to swell

**26. Cycle Length** - Length of the menstrual cycle of the female patient

**27. Baseline AMH** - Serum anti-Müllerian hormone (AMH) levels in female patient. AMH is created in the follicles and the concentration of AMH.

**28. Baseline FSH** - Follicle-Stimulating Hormone (FSH) levels in female patient. FSH stimulates the growth of follicles.

**29. Baseline LH** - Luteinizing Hormone (LH) levels in female patient. An increase in LH triggers the release of an egg during ovulation.

**30. Baseline E2** - Estradiol(E2) levels in female patient. E2 is a hormone produced within the ovaries

**31. DFI** - Sperm DNA Fragmentation Index in the male patient. High DFI indicates damage to the DNA of the sperm.

**32. Sperm Volume** - Sperm volume in male patients ejaculate.

**33. Sperm Concentration** - Sperm concentration in male patients ejaculate.

**34. Sperm Progressive Motility** - Percentage of progressively motile sperm in male patients ejaculate. Progressive motility indicates that the sperm moves in straight lines

**35. Sperm Non-Progressive Motility** - Percentage of non progressively motile sperm in male patients ejaculate.

**36. Sperm Motility** - Percentage of motile sperm in male patients ejaculate. Sperm motility is the sum of Non-Progressive and Progressive Sperm Motility.

**37. Sperm Amount** - Total sperm count in male patients ejaculate

**38. HDS** - High DNA Stainability, a sperm quality indicator.

**39. Fertilized Injected Eggs** - Number of fertilized eggs after ICSI (Intracytoplasmic sperm injection) treatment. During ICSI a single sperm is injected into the egg for fertilization to occur.

**40. Fertilized Inseminated Eggs** - Number of fertilized eggs after standard IVF treatment. During standard IVF treatment sperm and egg are placed together for fertilization to occur.

## 17.2 Oocytes: Forward selection

| Step | Added parameter | AIC after variable was added |
|---|---|---|
| 0 | Start model (Intercept) | 572.51 |
| 1 | Baseline AMH | 529.9 |
| 2 | Age | 525.31 |
| 3 | Baseline FSH | 521.84 |
| 4 | Indication | 518.92 |
| 5 | Baseline E2 | 516.49 |
| 6 | Baseline LH | 516.41 |

Table 32: Oocytes forward selection using AIC criterion

| Step | Added parameter | BIC after variable was added |
|---|---|---|
| 0 | Start model (Intercept) | 575.56 |
| 1 | Baseline AMH | 535.99 |
| 2 | Age | 534.36 |
| 3 | Baseline FSH | 534.04 |

Table 33: Oocytes forward selection using BIC criterion

## 17.3 Fertilization rate: Forward selection

| Step | Added parameter | AIC after variable was added |
|---|---|---|
| 0 | Start model (Intercept) | -413.05 |
| 1 | Age | -415.47 |
| 2 | Sperm Concentration | -416.02 |
| 3 | Indication | -419.12 |
| 4 | Oocytes | -419.77 |
| 5 | OSI | -420.42 |

Table 34: Fertilizaition rate forward selection using AIC criterion

| Step | Added parameter | BIC after variable was added |
|---|---|---|
| 0 | Start model (Intercept) | -410.03 |

Table 35: fertilization rate forward selection using BIC criterion

## 17.4 Fertilization group: Full LDA model weights and model selection

| Parameter | abs(Weight) |
|---|---|
| Fertilization method 2 | 2.066 |
| Fertilization method 3 | 1.381 |
| Indication 4 | 1.045 |
| Stim Days | 0.543 |
| Indication 2 | 0.473 |
| Indication 3 | 0.342 |
| Oocytes | 0.151 |
| BMI | 0.097 |
| OSI | 0.096 |
| Sperm Volume | 0.094 |
| Age | 0.087 |
| Dose per Day | 0.040 |
| Sperm Non-Progressive Motility | 0.028 |
| HDS | 0.017 |
| Sperm Progressive Motility | 0.016 |
| Baseline LH | 0.014 |
| Cycle Length | 0.010 |
| Progressive Sperm Amount | 0.008 |
| Progressive Sperm Amount | 0.008 |
| Baseline FSH | 0.006 |
| Baseline AMH | 0.005 |
| Sperm Amount | 0.005 |
| Total Dose | 0.003 |
| DFI | 0.003 |
| Sperm Concentration | 0.003 |
| Baseline E2 | 0.001 |

Table 36: LDA weights from LDA model with all parameters and Fertilization group as target variable

| # | Parameters | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|---|
| 0 | Naive model | 0.501 [0.404 0.607] | 0.534 [0.418 0.653] | 0.497 [0.367 0.625] | 0.507 [0.343 0.667] |
| 1 | 5 23 | 0.552 [0.472 0.629] | 0.669 [0.521 0.759] | 0.810 [0.451 0.980] | 0.215 [0.026 0.656] |
| 2 | 5 23 19 | 0.553 [0.466 0.636] | 0.671 [0.545 0.752] | 0.802 [0.500 0.978] | 0.223 [0.026 0.577] |
| 3 | 5 23 4 | 0.588 [0.505 0.663] | 0.691 [0.607 0.764] | 0.783 [0.635 0.918] | 0.380 [0.217 0.531] |
| 4 | 5 23 4 1 | 0.612 [0.523 0.693] | 0.700 [0.619 0.774] | 0.783 [0.635 0.918] | 0.340 [0.217 0.531] |
| 5 | 5 23 4 1 10 | 0.609 [0.523 0.693] | 0.698 [0.618 0.777] | 0.782 [0.640 0.913] | 0.376 [0.209 0.544] |
| 6 | 5 23 4 1 32 | 0.606 [0.511 0.693] | 0.694 [0.609 0.767] | 0.775 [0.622 0.915] | 0.377 [0.225 0.531] |
| 7 | 5 23 4 1 2 | 0.620 [0.523 0.693] | 0.702 [0.617 0.769] | 0.773 [0.618 0.911] | 0.415 [0.256 0.576] |
| 8 | 23 4 1 2 | 0.596 [0.511 0.682] | 0.691 [0.615 0.763] | 0.783 [0.618 0.935] | 0.344 [0.179 0.529] |
| 9 | 5 4 1 2 | 0.629 [0.552 0.708] | 0.720 [0.645 0.791] | 0.805 [0.644 0.933] | 0.377 [0.225 0.562] |
| 10 | 5 1 2 | 0.619 [0.531 0.698] | 0.716 [0.626 0.785] | 0.810 [0.613 0.961] | 0.346 [0.174 0.533] |
| 11 | 5 4 2 | 0.609 [0.526 0.691] | 0.709 [0.634 0.782] | 0.810 [0.619 0.954] | 0.324 [0.130 0.556] |
| 12 | 5 4 1 | 0.628 [0.552 0.708] | 0.724 [0.649 0.791] | 0.824 [0.656 0.960] | 0.347 [0.200 0.533] |

Table 37: Reduced LDA models with Fert Group as target variable.

## 17.5 GQE: Full LDA model weights and model selection

| Parameter | abs(Weight) |
|---|---|
| Fertilization Rate | 1.001 |
| Indication 2 | 0.851 |
| Indication 3 | 0.817 |
| Fertilization method 2 | 0.475 |
| Fertilization method 3 | 0.409 |
| Stim days | 0.310 |
| Sperm Volume | 0.285 |
| Fertilized injected eggs | 0.233 |
| Fertilized inseminated eggs | 0.207 |
| BMI | 0.150 |
| OSI | 0.099 |
| Baseline FSH | 0.067 |
| Age | 0.062 |
| Sperm Non-Progressive Motility | 0.047 |
| Oocytes | 0.046 |
| HDS | 0.036 |
| Baseline LH | 0.032 |
| Dose per Day | 0.032 |
| Progressive Sperm Amount | 0.026 |
| Indication 4 | 0.019 |
| Sperm Amount | 0.017 |
| DFI | 0.016 |
| Cycle Length | 0.016 |
| Sperm Progressive Motility | 0.010 |
| Baseline AMH | 0.004 |
| Sperm Concentration | 0.003 |
| Total Dose | 0.002 |
| Baseline E2 | 0.000 |
| Sperm Motility | 0.000 |

Table 38: LDA weights from LDA model with all parameters and GQE-group as target variable

| # | Parameters | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|---|
| 0 | Naive model | 0.497 [0.396 0.594] | 0.643 [0.545 0.734] | 0.495 [0.393 0.602] | 0.503 [0.095 1.000] |
| 1 | 6 23 | 0.639 [0.530 0.747] | 0.767 [0.672 0.847] | 0.648 [0.519 0.767] | 0.555 [0.167 1.000] |
| 2 | 6 23 5 | 0.640 [0.530 0.747] | 0.769 [0.672 0.847] | 0.654 [0.527 0.776] | 0.476[0.000 0.858] |
| 3 | 6 23 19 | 0.637[0.530 0.735] | 0.765 [0.672 0.844] | 0.646 [0.526 0.770 | 0.535 [0.142 1.000] |
| 4 | 6 23 32 | 0.653 [0.542 0.759] | 0.777 [0.693 0.851] | 0.662 [0.539 0.776] | 0.547 [0.167 1.000] |
| 5 | 6 23 32 39 | 0.645 [0.530 0.747] | 0.771 [0.672 0.847] | 0.655 [0.526 0.773] | 0.535 [0.143 1.000] |
| 6 | 6 23 32 40 | 0.650 [0.542 0.759] | 0.774 [0.683 0.855] | 0.657 [0.538 0.779 | 0.570 [0.167 1.000] |
| 7 | 6 23 32 1 | 0.687 [0.578 0.783] | 0.802 [0.720 0.873] | 0.692 [0.577 0.805] | 0.627 [0.200 1.000] |
| 8 | 6 23 32 1 10 | 0.690 [0.590 0.783] | 0.804 [0.721 0.875] | 0.696 [0.582 0.810] | 0.616 [0.200 1.000] |
| 9 | 6 23 32 1 10 28 | 0.680 [0.570 0.772] | 0.798 [0.708 0.868] | 0.692 [0.573 0.800] | 0.559 [0.167 1.000] |
| 10 | 6 23 32 1 10 2 | 0.689 [0.590 0.795] | 0.804 [0.726 0.878] | 0.697 [0.581 0.808] | 0.591 [0.200 1.000] |
| 11 | 23 32 1 10 | 0.651 [0.542 0.747] | 0.777 [0.689 0.851] | 0.666 [0.545 0.779] | 0.477 [0.000 0.857] |
| 12 | 6 32 1 10 | 0.661 [0.556 0.745] | 0.781 [0.697 0.848] | 0.663 [0.546 0.768] | 0.646 [0.286 1.000] |
| 13 | 6 23 1 10 | 0.681 [0.578 0.771] | 0.798 [0.715 0.868] | 0.687 [0.575 0.797] | 0.615 [0.200 1.000] |
| 14 | 6 23 32 10 | 0.662 [0.554 0.759] | 0.784 [0.699 0.859] | 0.673 [0.553 0.787] | 0.540 [0.143 1.000] |

Table 39: Reduced LDA models with GQE group as target variable.

## 17.6 Pregnancy: Full LDA model weights and model selection

| Parameter | abs(Weight) |
|---|---|
| Fertilization method 2 | 1.861 |
| Ferttilization method 3 | 1.493 |
| Indication 3 | 0.632 |
| GQE | 0.203 |
| Fertilized injected eggs | 0.202 |
| Indication 2 | 0.176 |
| ET Day | 0.137 |
| Fertilized inseminated eggs | 0.110 |
| Baseline FSH | 0.108 |
| Sperm Volume | 0.1041 |
| Cycle Length | 0.102 |
| BMI | 0.084 |
| Indication 4 | 0.068 |
| OSI | 0.033 |
| Oocytes | 0.030 |
| Age | 0.029 |
| Sperm Non-Progressive Motility | 0.024 |
| Stim Days | 0.023 |
| DFI | 0.022 |
| Sperm Progressive Motility | 0.018 |
| Progressive Sperm Amount | 0.0155 |
| Dose per day | 0.011 |
| Sperm Motility | 0.011 |
| Sperm Amount | 0.010 |
| Baseline AMH | 0.004 |
| HDS | 0.003 |
| Baseline LH | 0.001 |
| SP Concentration | 0.001 |
| Baseline E2 | 0.000 |
| Total Dose | 0.000 |

Table 40: LDA weights for the LDA model performed using all parameters that occur before Pregnancy on the timeline and tot-pregnant as target variable

| # | Parameters | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|---|
| 0 | Naive model | 0.501 [0.393 0.607] | 0.518 [0.386 0.638] | 0.496 [0.349 0.634] | 0.506 [0.349 0.667] |
| 1 | 5 23 | 0.552 [0.442 0.636] | 0.652 [0.482 0.738] | 0.795 [0.425 0.929] | 0.277 [0.105 0.531] |
| 2 | 5 23 8 | 0.542 [0.454 0.623] | 0.612 [0.474 0.710] | 0.692 [0.388 0.929] | 0.383 [0.154 0.724] |
| 3 | 5 23 8 39 | 0.545 [0.455 0.636] | 0.614 [0.480 0.714] | 0.689 [0.408 0.921] | 0.394 [0.171 0.697] |
| 4 | 5 23 8 39 7 | 0.544 [0.442 0.636] | 0.611 [0.482 0.710] | 0.681 [0.422 0.909] | 0.400 [0.189 0.667] |
| 5 | 5 23 8 39 40 | 0.544 [0.455 0.636] | 0.606 [0.494 0.706] | 0.664 [0.422 0.900] | 0.420 [0.194 0.657] |
| 6 | 5 23 8 39 40 28 | 0.541 [0.438 0.630] | 0.597 [0.474 0.695] | 0.642 [0.409 0.871] | 0.437 [0.216 0.700] |
| 7 | 5 23 8 39 40 32 | 0.535 [0.442 0.624] | 0.595 [0.481 0.697] | 0.648 [0.422 0.886] | 0.417 [0.210 0.647] |
| 8 | 5 23 8 39 40 26 | 0.544 [0.438 0.630] | 0.592 [0.472 0.690] | 0.628 [0.395 0.849] | 0.463 [0.243 0.697] |
| 9 | 5 23 8 39 40 1 | 0.565 [0.468 0.662] | 0.623 [0.506 0.713] | 0.677 [0.444 0.882] | 0.446 [0.245 0.656] |
| 10 | 23 8 39 40 1 | 0.533 [0.442 0.623] | 0.586 [0.475 0.681] | 0.624 [0.400 0.857] | 0.441 [0.225 0.667] |
| 11 | 5 8 39 40 1 | 0.545 [0.434 0.639] | 0.611 [0.488 0.699] | 0.668 [0.449 0.881] | 0.414 [0.178 0.657] |
| 12 | 5 23 39 40 1 | 0.565 [0.468 0.649] | 0.632 [0.524 0.722] | 0.702 [0.489 0.889] | 0.417 [0.220 0.618] |
| 13 | 5 23 40 1 | 0.550 [0.455 0.636] | 0.625 [0.512 0.713] | 0.705 [0.475 0.892] | 0.380 [0.182 0.581] |
| 14 | 5 23 39 1 | 0.562 [0.494 0.706] | 0.639 [0.538 0.717] | 0.724 [0.537 0.898] | 0.384 [0.187 0.600] |
| 15 | 5 23 39 | 0.555 [0.455 0.649] | 0.646 [0.459 0.733] | 0.769 [0.367 0.925] | 0.315 [0.150 0.594] |

Table 41: Reduced LDA models with Tot-Pregnant as target variable.

## 17.7 Live Birth: Full LDA model weights and model selection

| Parameter | abs(Weight) |
|---|---|
| Stim Days | 0.243 |
| Fertilized inseminated eggs | 0.208 |
| BMI | 0.159 |
| ET Day | 0.156 |
| GQE | 0.154 |
| OSI | 0.139 |
| Fertilized injected eggs | 0.137 |
| Cycle Length | 0.131 |
| Sperm Volume | 0.084 |
| Non-progressive Motility | 0.033 |
| Progressive sperm amount | 0.021 |
| Baseline AMH | 0.021 |
| Baseline LH | 0.015 |
| HDS | 0.014 |
| Baseline FSH | 0.012 |
| Dose per day | 0.012 |
| Sperm amount | 0.011 |
| Oocytes | 0.011 |
| Age | 0.009 |
| Progressive Motility | 0.005 |
| Sperm Concentration | 0.005 |
| Total dose | 0.001 |
| DFI | 0.001 |
| Motility | 0.001 |
| Baseline E2 | 0.001 |

Table 42: LDA weights for the LDA model performed using all parameters that occur before Live Birth on the timeline and Total Live Births as target variable

| # | Parameters | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|---|
| 0 | Naive model | 0.498 [0.348 0.630] | 0.605 [0.456 0.746] | 0.497 [0.343 0.658] | 0.501 [0.143 0.800] |
| 1 | 19 40 | 0.471 [0.333 0.600] | 0.577 [0.400 0.725] | 0.473 [0.281 0.676] | 0.468 [0.100 0.800] |
| 2 | 19 40 1 | 0.528 [0.400 0.667] | 0.630 [0.478 0.769] | 0.524 [0.344 0.711] | 0.543 [0.250 0.857] |
| 3 | 19 40 1 7 | 0.524 [0.378 0.645] | 0.624 [0.462 0.758] | 0.516 [0.333 0.703] | 0.552 [0.250 0.875] |
| 4 | 19 40 1 8 | 0.550 [0.400 0.667] | 0.644 [0.489 0.769] | 0.530 [0.351 0.703] | 0.626 [0.300 0.917] |
| 5 | 19 40 1 8 10 | 0.547 [0.422 0.689] | 0.644 [0.480 0.772] | 0.533 [0.353 0.706] | 0.601 [0.250 0.900] |
| 6 | 19 40 1 8 39 | 0.547 [0.400 0.689] | 0.644 [0.481 0.769] | 0.532 [0.343 0.710] | 0.603 [0.250 0.900] |
| 7 | 19 40 1 8 26 | 0.554 [0.415 0.707] | 0.654 [0.476 0.793] | 0.550 [0.345 0.742] | 0.569 [0.167 0.889] |
| 8 | 19 40 1 8 26 32 | 0.554 [0.415 0.707] | 0.659 [0.500 0.793] | 0.563 [0.375 0.765] | 0.521 [0.143 0.875] |
| 9 | 19 1 8 26 | 0.563 [0.429 0.690] | 0.661 [0.491 0.788] | 0.555 [0.367 0.742] | 0.592 [0.200 0.900] |
| 10 | 1 8 26 | 0.564 [0.429 0.690] | 0.658 [0.489 0.788] | 0.550 [0.357 0.742] | 0.613 [0.200 1.000] |
| 11 | 8 26 | 0.548 [0.381 0.690] | 0.633 [0.419 0.783] | 0.516 [0.281 0.750] | 0.669 [0.167 1.000] |
| 12 | 1 26 | 0.524 [0.381 0.690] | 0.630 [0.458 0.793] | 0.532 [0.333 0.758] | 0.495 [0.143 0.800] |
| 13 | 1 8 | 0.561 [0.435 0.696] | 0.651 [0.489 0.781] | 0.530 [0.341 0.714] | 0.676 [0.300 1.000] |

Table 43: Reduced LDA models with Live Birth as target variable.