

Document downloaded from:

<http://hdl.handle.net/10251/49738>

This paper must be cited as:

Debón Aucejo, AM.; Molina Botella, MI.; Cabrera García, S.; Pellicer, A. (2013).
Mathematical methodology to obtain and compare different embryo scores. *Mathematical and Computer Modelling*. 57(5-6):1380-1394. doi:doi:10.1016/j.mcm.2012.11.027.



The final publication is available at

<http://www.sciencedirect.com/science/article/pii/S0895717712003408>

Copyright Elsevier

Mathematical methodology to obtain and compare different embryo scores

A. Debón^{a,e}, I. Molina^{b,c}, S. Cabrera^d, A. Pellicer^c

^a*Centro de Gestión de la Calidad y del Cambio
Universitat Politècnica de València.*

E-46022. Valencia. Spain.

^b*Dpto. de Ciencia Animal*

Universitat Politècnica de València.

E-46022. Valencia. Spain.

^c*Unidad de Reproducción Humana*

Hospital Universitario La Fe de Valencia

E-46009. Valencia. Spain.

^d*Grupo de Fenómenos Ondulatorios*

Universitat Politècnica de València.

E-46022. Valencia. Spain.

^e*Corresponding author. Tel: +34 963877007 (Ext. 74961); Fax: +34 963877499.*

E-mail: andeau@eio.upv.es

Abstract

In Vitro Fertilization (IVF) units need to decrease multiple pregnancies without affecting their overall success rate. In this study we propose a mathematical model to evaluate an embryo's potential ability to implant in the uterus. Embryos are graded by the embryologist based on the number of blastomeres, evenness of growth and degree of fragmentation. Therefore, the following variables were considered: number of blastomeres produced by division of the egg after fertilization (blastomeres), symmetry and fragmentation of the embryo (grade). This model evaluates the embryos assigning them a

score which represents their quality. The main result derived from this model is the estimation of the significant improvement in the implantation rate due to the increase in blastomere values and the decrease in grade factor values. But the increase from two-three to four produces more improvement in the implantation rate than two-three to five-six blastomeres.

First, statistical models were used to study embryo traceability from transfer to implantation and to evaluate the effect of the quality of the embryos (embryo score) and women's age on implantation potential. This score was obtained by making predictions from the fitted model which was used to rank embryos in terms of implantation potential. Then we totalled the scores of embryos that had been transferred to each woman for obtaining the Embryo Quality Index (EQI). In addition, we studied the effects of EQI and women's age on pregnancy. Finally, statistical techniques such as Receiver Operating Characteristics (ROC) and bootstrap procedures were used to assess the accuracy of this model. This embryo score is a quick, efficient and accurate tool to optimize embryo selection for transfers on the second day after fertilization. This tool is especially useful for transfers involving non-top embryos.

Keywords: Generalised linear models, Categorical predictors, Generalised

linear mixed models, ROC curve, Embryo score, Bootstrapping.

1. Introduction

The increased obstetric and perinatal risks involved in multiple pregnancies urge the clinician to reduce the number of embryos to transfer following in vitro fertilization (IVF)-intracytoplasmic sperm injection (ICSI) cycles [1–3]. IVF and ICSI assisted reproduction techniques are highly complex. they are so called because it is necessary to have an embryology laboratory for the handling of gametes and embryos in vitro, ie outside the woman’s body. Embryo transfer refers to a step in reproduction techniques in which embryos are placed in the uterus of a female with the intent of establishing a pregnancy (embryo implantation). High multiple pregnancy rates correlate both to the number and quality of transferred embryos, therefore [3–5] propose to select only one top quality embryo to transfer (SET). The implementation of a top quality single-embryo transfer (SET), should produce an important decrease in multiple pregnancies without a significant decrease in pregnancy rates [6–10]. However, the implementation of SET determines an unacceptably low pregnancy rate, particularly in older women and in those with poor embryo quality [10]. Thus, it is important to increase our knowledge of the implantation potential of each individual embryo in order to select top quality

embryos for transfer. At the same time a ranking selection, which allows the reduction of multiple pregnancies without reducing pregnancy rates, should be established.

The two factors that most influence the implantation rate are a woman's age and embryo quality [11–18]. A woman's age is unchangeable. However, when there is a sufficient number of embryos available, we can select the embryos to be transferred with the greatest implantation potential according to morphological criteria such as blastomere number and blastomere symmetry, equality and multinuclearity [14, 18–25]. Evaluation of the implantation potential of transferred embryos has generally been based on the construction of accumulated embryonic scores. Assumptions need to be made about the overall quality of transferred embryos and their subsequent implantation due to a lack of knowledge about of the exact quality of the embryo finally implanted [11, 13, 16, 26–30]. In more recent studies, logistic regression models have been used to predict the possibility of embryo implantation [18]. The inclusion of embryonic quality as continuous instead of categorical variables or factors in the logistic regression models, forces those authors to use transformed variables. It also does not provide knowledge about which values of the variable produce significant increases in the implantation rate.

In the present study distinctive methodology for estimating the embryos' implantation potential has been developed. This methodology allows us to evaluate the effect of each variable's value when it is considered as a factor. In addition, we propose the validation of models by using Receiver Operating Characteristics (ROC) curves. The aim of this paper is to propose a mathematical methodology to obtain and compare different embryo scores adapted to the number and nature of our database variables.

2. Materials and Methods

2.1. Data

The paper is a retrospective study of 5242 cycles of IVF-ICSI with transfers of one, two or three embryos on day 2 (second day after fertilization) in the Human Reproduction Unit at the University Hospital La Fe in Valencia from January 2003 to January 2007.

2.2. IVF-ICSI procedure

The women were treated using a controlled ovarian hyper-stimulation protocol (COH), including down regulation with a gonadotropin-releasing hormone (GnRH) agonist in a long protocol. Stimulation was performed with recombinant follicle stimulating hormone (FSH: Gonal or Puregon).

Oocytes were retrieved 36-38 h following human chorionic gonadotrophin (HCG) using transvaginal sonographically guided puncture.

Fertilization was performed by conventional IVF or ICSI, following standard techniques [31]. A commercial culture media was used according to local routines (MediCult Denmark). The oocytes were inseminated (or injected with sperm after denudation for ICSI) after 2-6 h of incubation and cultured in an IVF medium (MediCult, Denmark) in a 5% CO₂ incubator at 37°C. All ICSI were performed with motile spermatozoa. When both ICSI and IVF embryos were available, the best quality embryos were transferred. Fertilization was checked 16-20 h after insemination. The embryos were evaluated and transferred on day 2 after oocyte retrieval. Selection was performed immediately before embryo transfer.

The embryos were selected depending on cleavage rate and blastomere symmetry size and fragmentation. The embryo classification was modified from the system described by [28] as follows:

- Grade 1 (G1) embryos consisted of symmetrical blastomeres of approximately equal size and without anucleate fragments.
- Grade 2 (G2) embryos had blastomeres of even or uneven size and had less than 15% of the volume of embryos filled with anucleate fragments.

- Grade 3 (G3) embryos had anucleate fragments occupying between 15 and 50% of the volume of the embryos.
- Grade 4 (G4) embryos had anucleate fragments occupying more than 50% of the volume of the embryos.

Then, all the transferred embryos were scored considering the following variables: the number of blastomeres (2, 3, 4, 5 and 6) and the grade of fragmentation and variation in the size of the blastomeres (Grades 1, 2, 3 and 4). To evaluate both the number of blastomeres and fragmentation grade of the embryos, the cumulative embryo score (CES) [11] for each embryo transfer was calculated. The CES was calculated multiplying the number of blastomeres from each embryo by their grade recoded numerically as G1 = 4, G2 = 3, G3 = 2 and G4 = 1. Thus the embryos with the highest score were those which had a higher number of blastomeres and less degree. It is a criterion used to select higher quality embryos for later transfer.

Embryo transfer was sonographically controlled. Two selected embryos, depending on their availability, were transferred and only 3 embryos were transferred in those cases with inadequate embryo quality and/or higher female age (more than 37 years old). Luteal support was given with vaginal progesterone.

Pregnancy was defined as a gestational sac with a foetus with heart activity detected through sonography in gestational week 7-8. The implantation rate was defined as the number of gestational sacs per number of transferred embryos, and double and triple pregnancy as the number of double or triple sacs per number of pregnancies.

The exact fate or implantation potential of each transferred embryo could only be traced in: single pregnancies of one transferred embryo, twin pregnancies of two transferred embryos, triple pregnancies of three transferred embryos or treatments with a negative pregnancy test. In this study we considered a total number of 5242 cycles and the corresponding 11362 embryos distributed in the following groups:

- Group 0: 3577 cycles with transfers of one, two or three embryos which gave a negative pregnancy test corresponding to 7489 embryos.
- Group 1: 61 cycles with only one transferred embryo resulting in a single pregnancy (only one gestational sac), therefore 61 embryos.
- Group 2: 213 cycles with two transferred embryos resulting in twin pregnancies (two gestational sacs), therefore 426 embryos.
- Group 3: 52 cycles with three transferred embryos resulting in triplet

pregnancies (three gestational sacs), therefore 156 embryos.

- Group 4: 1339 transfers with a lower number of ultrasonically identified sacs than transferred embryos. This last group included transfers of two embryos with only one gestational sac as well as transfers of three embryos with both, two and only one gestational sac, the rest of the embryos 3230.

In the statistical analysis, each embryo was treated as an individual record. Thus, 643 embryos with a correct implantation, i.e. Groups 1, 2 or 3 and 7489 with a proven failed implantation, i.e. Group 0, were used to estimate the model. Once the model had been established, we assigned a score to each type of embryo, from which we totalled the scores of embryos that had been transferred to each woman i.e. the Embryo Quality Index (EQI). The EQI was validated with the data of all the women in the study.

2.3. Statistics

A statistical analysis was performed using the R environment for statistical computing [32]. The ROCR package [33] was also used. ROCR is a package for evaluating and visualizing the performance of scoring classifiers using the statistical language R. This package makes it easy to use a Receiver

Operating Characteristics (ROC) graph as a way to evaluate concordance between models and real data.

In this study we used a Generalised Linear Model (GLM, [34]) to calculate the predictive value of the categorical variables, such as the number of blastomeres and the grade, for the occurrence of ongoing pregnancy. A $p\text{-value} < 0.10$ was considered statistically significant. GLM allowed us to analyze binary data and logit models, with categorical predictors often called factors. A detailed description is available in [35]. The coefficients in logit models with categorical variables are used to study the differences in probabilities between different values in the independent variables. However, one of the underlying assumptions of this approach (GLM) is that the data are independent, which is not always the case. In this paper, we take this into account by using Generalised Linear Mixed Models (GLMM), which extend the GLM models to allow for correlation between the observations and nested data structures. These models are fully described in [36]. There are various packages in R that can be used for GLMM. We used the `lme4` package [37] because it makes model comparison easier by providing an Akaike information criterion (AIC). AIC is a measure of the relative goodness-of-fit of a statistical model. Hence, AIC not only rewards goodness-of-fit, but also

includes a penalty that is an increasing function of the number of estimated parameters.

ROC curves provided an overall representation of accuracy, and they are well-described by [38]. If the test did not allow discrimination between classes, the ROC curve was the diagonal joining the vertices from lower left to upper right. The accuracy of the test increased as the curve moved towards the upper left corner.

Our model was designed to show the impact of each value of categorical variables for the implantation potential of the individual embryo on day 2 of transfer. To evaluate the discriminative performance of the logistic model and to compare the classifiers, we wanted to reduce ROC performance to a single scalar value representing expected performance. Calculating the area under the ROC curve of the classifier, in short *AUC*, is a common method. Since the *AUC* was a portion of the area of the square unit, its value was always between 0 and 1, so random guessing procedures had an area of 0.5. Therefore, when the area under the ROC curve (*AUC*) increased, the classifier power also increased.

To verify whether the differences between the *AUC* of the ROC curves for the models compared are significant we used bootstrapping procedures

[39], in the following way: starting from the observations, we selected N bootstrap samples with n observations each. For each bootstrap sample, the AUC s were computed for each model. This yielded N realizations which were used to compute,

1. The bootstrap distribution of the AUC s and the bootstrap distribution of the difference of the AUC s,
2. the corresponding confidence intervals by means of percentiles 2.5 and 97.5, and
3. the statistic to compare the ROC curves based on their AUC s.

Comparing ROC curves with AUC statistics [40], two curves are denoted ROC_A and ROC_B , where the null hypothesis is:

$$H_0 : ROC_A = ROC_B,$$

based on the comparison

$$H_0 : AUC_A = AUC_B$$

though we really solve

$$H_0 : \text{mean}(AUC_A) = \text{mean}(AUC_B)$$

with paired the t-test as Hanley and McNeil [41] warn that the comparison is carried out on the same sample. The normality of the $(AUC_A - AUC_B)$ was tested through QQ-plot.

The last models used in this paper were the generalized additive model (GAM). The GAM is a statistical model developed by Hastie and Tibshirani [42] for blending properties of generalized linear models with additive models. GAMs and GLMs can be applied in similar situations, but they serve different analytic purposes. GLMs emphasize estimation and inference for the parameters of the model, while generalized additive models focus on exploring data nonparametrically. GAMs are more suitable for exploring the data set and visualizing the relationship between the dependent variable and the independent variables. Therefore GAMs can be used as a prognostic tool for the investigation of logistic regression.

In addition, and first of all, we are going to use Correspondence Analysis (CA), which is a multivariate statistical technique conceptually similar to Principal Component Analysis (PCA), but applies to categorical rather than continuous data. In a similar manner to PCA, it provides a means of displaying or summarising a set of data in two-dimensional graphical form. Recently, several related R packages have implemented this technique. For

instance, the `anacor` package by de Leeuw and Mair [43] offers additional possibilities for scaling the scores in simple CA and canonical CA.

3. Results

The total clinical pregnancy rate was 31.76% (total number of pregnancies/total number of cycles = 1665/5242), with a twin rate of 22.22% (370/1665), and a triplet rate of 3.17 % (53/1665), corresponding to an implantation rate of 17.37% (total number of sacs/total number of transferred embryos =1974/11362). When only considering embryos with correct implantation, this rate was reduced to 7.89% (total number of embryos with correct implantation/total number of embryos with correct or failed implantation =643/8146).

3.1. Building the model: Impact of morphological variables on the implantation: analysis restricted to Groups 0, 1, 2 and 3

The basic aim of our analysis is to predict the way in which implantation potential varies by embryonic characteristics (number of blastomeres and grade) and therefore it is important to note if the predictors are independent discrete factors. We can use the Bernoulli distribution for binary response variable “correct” or “failed” *implantation* (i.e., 1 or 0) or the binomial distribution for grouped data. Therefore, we were able to fit a logit

model to predict implantation potential of each individual embryo, from the categorical predictors: number of **blastomeres** and **grade** for embryos from Groups 0,1,2,3. The number of successes, total of transferred embryos and the corresponding crude implantation rates are shown in Tables 1, 2 and 3 respectively.

	Blastomeres					
Grade	2	3	4	5	6	
1	29	3	178	4	2	
2	35	9	293	10	8	
3	10	5	54	2	0	
4	1	0	0	0	0	

Table 1: Number of corrected implanted embryos at different levels for the two embryo variables in groups 1, 2 and 3.

Since there were so few observations in some categories such as 3, 5 or 6 blastomere values with low implantation rates (Table 2 and Table 3), as a first step levels were combined based on CA. The CA was obtained using the `anacor` function in the `anacor` package for R. The following code was used in R:

Blastomeres					
Grade	2	3	4	5	6
1	584	95	1329	20	29
2	835	320	2463	96	146
3	512	198	1218	60	60
4	65	27	66	5	4

Table 2: Total of transferred embryos at different levels for the two embryo variables in groups 0, 1, 2 and 3.

Blastomeres					
Grade	2	3	4	5	6
1	4.97	3.16	13.39	20.00	6.90
2	4.19	2.81	11.90	10.42	5.48
3	1.95	2.53	4.43	3.33	0.00
4	1.54	0.00	0.00	0.00	0.00

Table 3: Crude implantation rates (%) at different levels for the two embryo variables in groups 0, 1, 2 and 3.

```
correct <- c(29,3,178,4,2,35,9,293,10,8,10,5,54,2,0,1,0,0,0,0)
```

```
total <- c(584,95,1329,20,29,835,320,2463,96,146,512,198,1218,
```

```

60,60,65,27,66,5,4)

rates<-correct/total

rates[rates==0]<-0.01

table<-matrix(ratios,nrow=4,ncol=5,byrow=TRUE,

              dimnames=list(c("grade1","grade2","grade3","grade4")

                             ,c("blast2","blast3","blast4","blast5","blast6")))

library(anacor)

res<-anacor(table, scaling=c("Goodman","centroid"))

plot(res, plot.type="jointplot")

```

The CA plot is shown in Figure 1 where the distances within grade categories are quite large, but not within the number of blastomeres. We see that 2 and 3 as well as 5 and 6 are quite close to each other. Therefore, categories for the number of blastomeres were grouped into 2-3, 4 and 5-6.

The saturated model (main effects plus interaction) was fitted using the `glm` function in the `stats` package for R. The following code was used in R:

```

correct.agrup <- c(32,178,6,44,293,18,15,54,2,1,0,0)

total.agrup <- c(679,1329,49,1155,2463,242,710,1218,120,92,66,9)

embryo.agrup <-data.frame(grade=rep(1:4, each=3),

blastomeres=rep(c(2,4,6),4), correct.agrup,total.agrup)

```

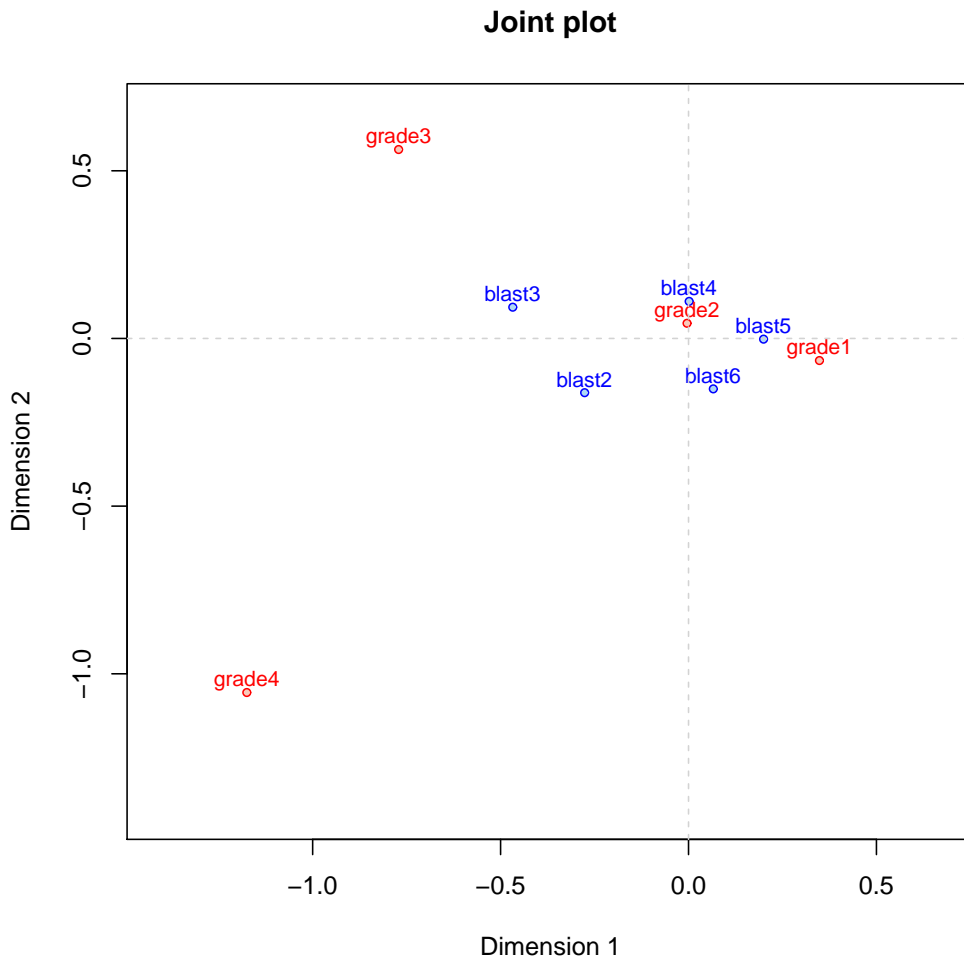


Figure 1: Result of the Correspondence Analysis

```

Model.glm<-glm(cbind(correct.agrup,total.agrup-correct.agrup)~
factor(blastomeres)+factor(grade)+factor(blastomeres):factor(grade),
family=binomial(link=logit)),

```

the `factor` command is used to indicate which independent variables are categorical. The next step is to select the most relevant effects for adjustment. The R command `step` is used to select a formula-based model by AIC (the preferred model is the one with the lowest AIC value). This command chooses the minimum number of significant effects that fit best, in this case the model with only main effects and without interaction. The results of this logit model are shown in Table 4, which includes the value of the coefficients for each of the variables values, Estimate, the standard error (SE), the t-value and the p-value or significance for each of the coefficients. Parameters for the reference values, 2 and 1, number of blastomeres and grade, respectively, do not appear in this table.

The interpretation of the positive coefficient β which corresponds to the number of blastomeres (indicated as a subindex of the parameter) can be interpreted as that the logit implantation rate is higher in the embryos with this value than in the reference value. On the contrary, the interpretation of the negative coefficient is that the logit implantation rate is lower in this value than in the reference value. This is true if we consider the grade effect which corresponds to γ parameters.

From the results of the adjusted model in Table 4 we were able that the

Parameter	Estimate	SE	t-value	p-value ^a
α	-2.9824	0.12099	-24.650	0.0000***
β_4	1.1222	0.11616	9.661	0.0000***
β_{5-6}	0.6319	0.23027	2.744	0.0061**
γ_2	-0.1581	0.09198	-1.719	0.0856
γ_3	-1.1636	0.14136	-8.231	0.0000***
γ_4	-2.7518	1.00625	-2.735	0.0062**

^aSignif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 4: Estimates for the parameters of logit model with categorical predictors.

number of blastomeres variable increase the logit of the implantation rate, specifically, the increase of 2-3 to 4 blastomeres increased the logit by 1.1222, while the 5-6 blastomere embryos increased the logit by 0.6319. The grade variable, however, decreased this logit as moving from G1 to G2, G3 or G4 produced a decrease of 0.1581, 1.1636 and 2.7518 respectively.

The implantation rate logit, $\text{logit}(p) = a$, was estimated from this model adding the corresponding parameters to constant, α which is the implantation rate logit for reference embryos with number of blastomeres 2 and grade 1. Therefore, the corresponding implantation rate was obtained from the

estimated rate logit, a , by using its inverse function,

$$p = \exp(a)/(1 + \exp(a)). \quad (1)$$

Table 5 shows the percentage of implantation rates corresponding to each combination of number of blastomeres and grade for each of the embryos considered. The low values of the previous implantation rates were due to the fact that we had only considered embryos with a correct implantation in the numerator. In any case the results in Table 5 show the quality of the embryos evaluated and we can use them as a score for each embryo.

	blastomeres		
Grade	2-3	4	5-6
1	4.82	13.47	8.70
2	4.15	11.73	7.52
3	1.56	4.63	2.89
4	0.32	0.98	0.60

Table 5: Implantation rates (%) for logistic model with categorical predictors for all possible combinations of blastomere number and grade.

However, embryos from the same mother are clearly dependent and this ought to have been accounted for in earlier methods. The point is that when

two or three embryos from the same mother are implanted, other important (unmeasured) information on the mother-level is equal for these embryos, increasing the likelihood of having the same outcome. This was accounted for by using GLMM with women as a random effect which implies that the probability of embryo implantation is correlated with other embryos for the same mother. In order to check the consistency of these previous results, the GLMM model was fitted using the `lme` function in the `lme4` package for R but using the binary variable *implantation* for each embryo. The following code was used in R:

```
library(lme4)

embrion$celagrup[embrion$blastomeres==2 | embrion$blastomeres==3]<-2
embrion$celagrup[embrion$blastomeres==5 | embrion$blastomeres==6]<-6

Model.glmm<-glmer(implantation~factor(celagrup)+factor(grade)
+factor(celagrup):factor(grade)+(1|historia),
  family=binomial(link=logit), data=embrion,nAGQ=5)
```

Again, in this case the interaction does not produce a significant effect on the implantation rate. The significant factors have not changed from the above Table 4, and the results of this mixed model are shown in Table 6.

Finally, we totalled the scores of embryos (Table 5) that had been trans-

Parameter	Estimate	SE	t-value	p-value ^a
α	-5.9472	0.3111	-19.118	0.0000***
β_4	0.9414	0.2780	3.386	0.0007***
β_{5-6}	0.2877	0.5978	0.481	0.6304
γ_2	-0.4068	0.2162	-1.881	0.0599
γ_3	-1.5705	0.3521	-4.461	0.0000***
γ_4	-3.5376	2.2962	-1.541	0.1234

^aSignif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 6: Estimates for the parameters of mixed logit model with categorical predictors.

ferred to each woman or cycle and we studied the effect of women’s age on the pregnancy, in the case of the database being for **women**.

3.2. Validation of the model: the discriminating power of the model for each woman

Receiver Operating Characteristics (ROC) curves and the corresponding *AUC* were used to validate our model. Then, we considered the discriminating power of the score for each embryo proposed in Table 5 whose implantation potential varies according to embryonic characteristics (number of blastomeres and grade) in order to select the best embryos, thereby mak-

ing a quick decision. After, we totalled the scores for embryos that were transferred to every woman who obtained a minimum EQI between 0.32 (one embryo with 2-3 blastomeres and grade 4) and 40.26 (tree embryo with 4 blastomeres and grade 1). In this study the *AUC* was 0.6698 for EQI. The study validates the model by taking into account the embryos from the same mother, obtaining EQI in order to predict pregnancy. In order to compare our model with those used by other authors we calculated the CES proposed by Steer et al. [11] for our data and the corresponding *AUC* was 0.6564, which was lower than our value. We also compared our results with a logistic model with continuous predictors (LMCP) based on proposal [18]. Those authors constructed a model using blastomere equality and symmetry variables independently, at the same time including the blastomere multinucleation variable. However, our database did not allow us to consider the first two variables independently. Therefore they were evaluated together as the grade variable. In our study, the multinucleation variable was not taken into account.

This model was fitted using the same function `glm` but the R code was:

```
correct <- c(29,3,178,4,2,35,9,293,10,8,10,5,54,2,0,1,0,0,0,0)
total <- c(584,95,1329,20,29,835,320,2463,96,146,512,198,1218,
```

```

60,60,65,27,66,5,4)

embryo<-data.frame(grade=rep(1:4, each=5),blastomeres=rep(2:6,4),
correct,total)

Model.LMCP<-glm(cbind(correct,total-correct)~ blastomeres
+ abs(blastomeres-4) + grade, family=binomial(link=logit)),

```

without the factor command as variables are continuous, therefore, values for the number of blastomeres were 2, 3, 4, 5 and 6. In this model all the variable coefficients are significant. We used the LMCP and the transformations of blastomeres suggested by [18]. In this case, the implantation rate logit was estimated from the regression model $\text{logit}(p) = -1.7319 + 0.1323 * \text{blastomeres} - 0.4689 * \text{abs}(\text{blastomeres} - 4) - 0.5125 * \text{grade}$. Therefore, the corresponding implantation rate was obtained, using the inverse function (1) as previously indicated. The results of implantation rate, expressed as a percentage, corresponding to each combination of number of blastomeres and grade are shown in Table 7.

The *AUC*s corresponding to this methodology was 0.6664, very similar to that obtained with our model, although slightly lower.

When comparing results, the *AUC* values are very similar. For that reason, we obtained the statistics for 1000 bootstrap samples, Figure 2 shows the

blastomeres					
Grade	2	3	4	5	6
1	5.13	8.98	15.25	11.39	8.41
2	3.14	5.58	9.73	7.15	5.21
3	1.90	3.42	6.07	4.41	3.19
4	1.15	2.08	3.72	2.69	1.93

Table 7: Implantation rates (%) for logistic model with continuous predictors (LMCP) for all possible combinations of blastomere number and grade.

bootstrap distribution of the AUCs and Figure 3 the bootstrap distribution of the difference of the AUCs.

To compare the embryo scores obtained before we calculated confidence intervals (CI) for *AUCs* and their differences. In addition, we applied a paired t-test evaluated with EQI as the model A and each of LMCP and CES as model B. Table 8 shows the results. According to these t-values and their corresponding p-values it can be concluded that the “true” mean of *AUCs* of EQI are greater than the “true” mean of *AUCs* of the LMCP and CES.

The normality of the difference of the AUCs for EQI vs LMCP and EQI

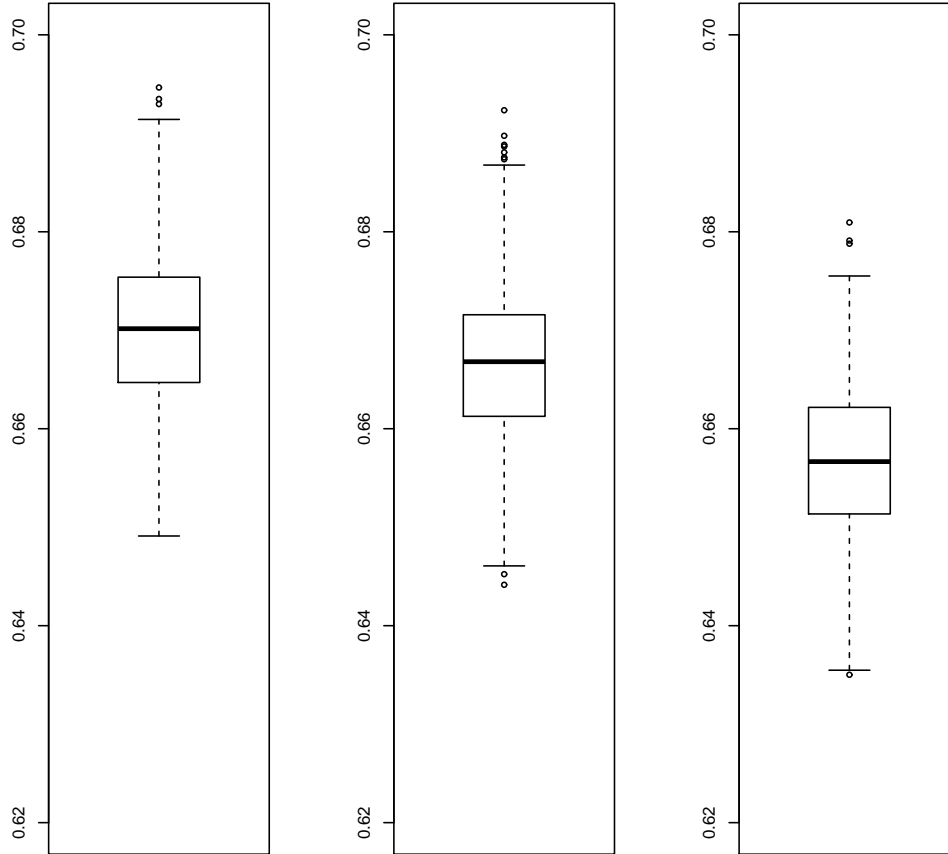


Figure 2: The bootstrap distribution of the AUC for EQI, LMCP and Steer

vs CES were tested through QQ-plot using the `qqnorm` function in the `stats` package for R. Figure 4 shows the resulting QQ-plot showing approximately an straight line that supports the normality and consequently the corrected

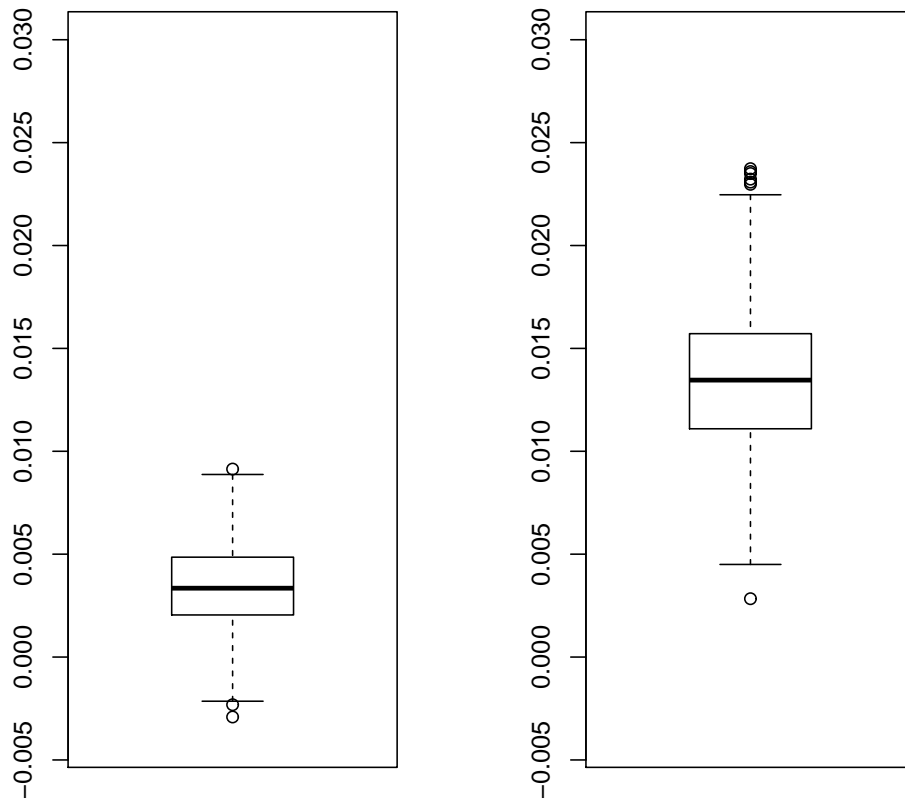


Figure 3: The bootstrap distribution of the difference of the AUCs for EQI vs LMCP and EQI vs CES

application of the paired t-test.

Figure 5 illustrates the ROC curves for the three models whose compari-

Model	CI(95%)	CI(95%)	t-value	p-value
EQI	[0.6555,0.6820]			
CES	[0.6424,0.6715]	[0.0069,0.0205]	124.04	0
LMCP	[0.6521,0.6821]	[-0.0008,0.0074]	52.09	0

Table 8: The confidence intervals (CI) for AUC's and their differences and the results of t-test

son allows us to assert that our model assigns scores that discriminate better between women who are pregnant or not as the other two models provide curves closer to the diagonal.

In Figure 5 we have marked the point nearest the upper left corner on each curve which would be the best-cut decision, also called the optimal operating point. These cut-offs are 24, 18.93 and 19.08 for CES, LMCP and EQI respectively. These values correspond to the cut-off points we must make in the scores of the models assigned to each woman in order to differentiate between pregnant and non-pregnant women so that the classification can be as near as possible to the observed values.

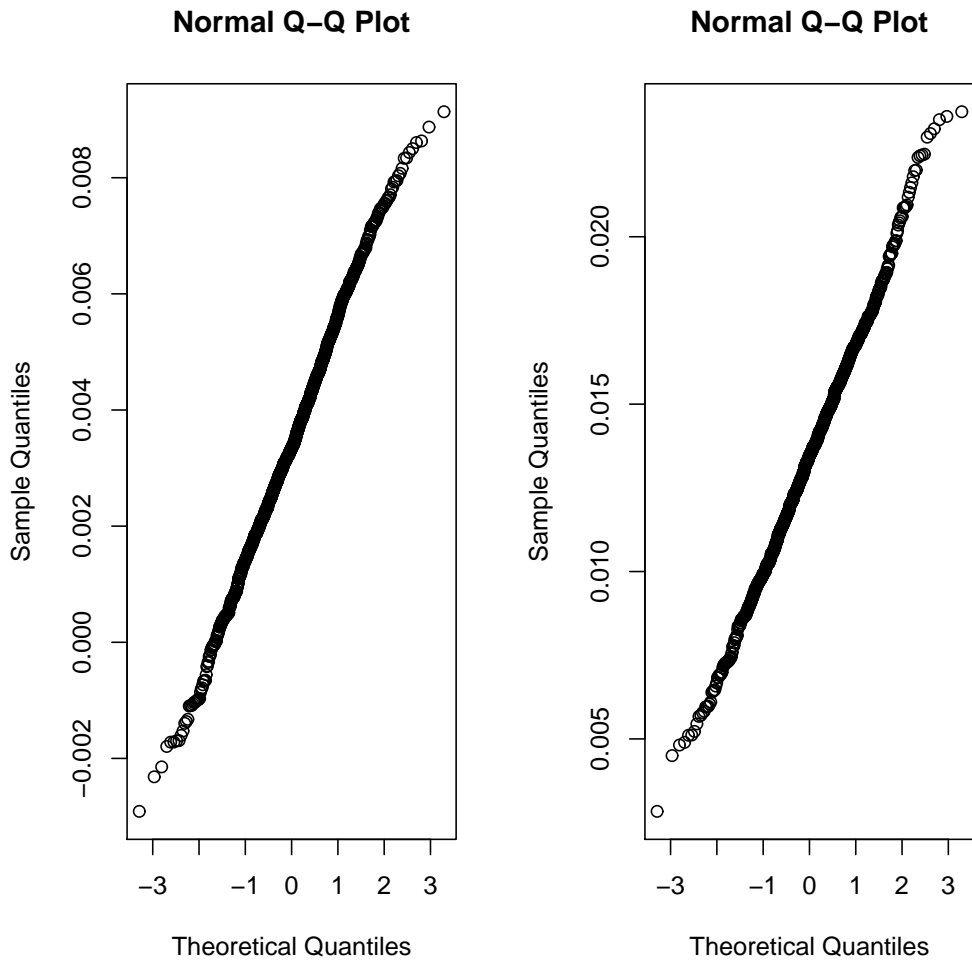


Figure 4: The QQ plot of the difference of the AUCs for EQI vs LMCP and EQI vs CES

3.3. Validation of the model: the discriminating power of the model for each embryo

Finally, 75% of the embryo data were randomly selected to establish the models and the remaining 25% were saved exclusively for evaluation. We

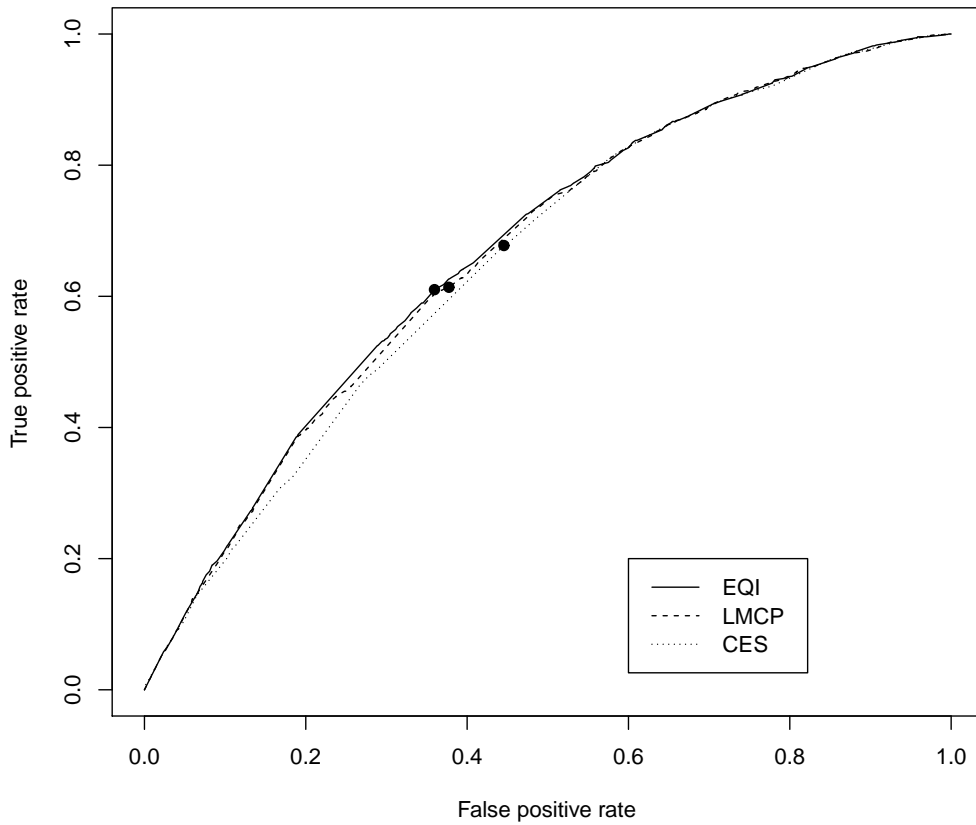


Figure 5: Comparison of the ROC curves for women

repeated this process obtaining 1000 different 75% embryo data samples for modeling and the remaining 25% for validation. We obtained the AUCs corresponding to the 1000 bootstrap samples with 25 % of the embryo data for validation. Figure 6 shows the bootstrap distribution of these AUCs and

Figure 7 the bootstrap distribution of the difference between the AUCs.

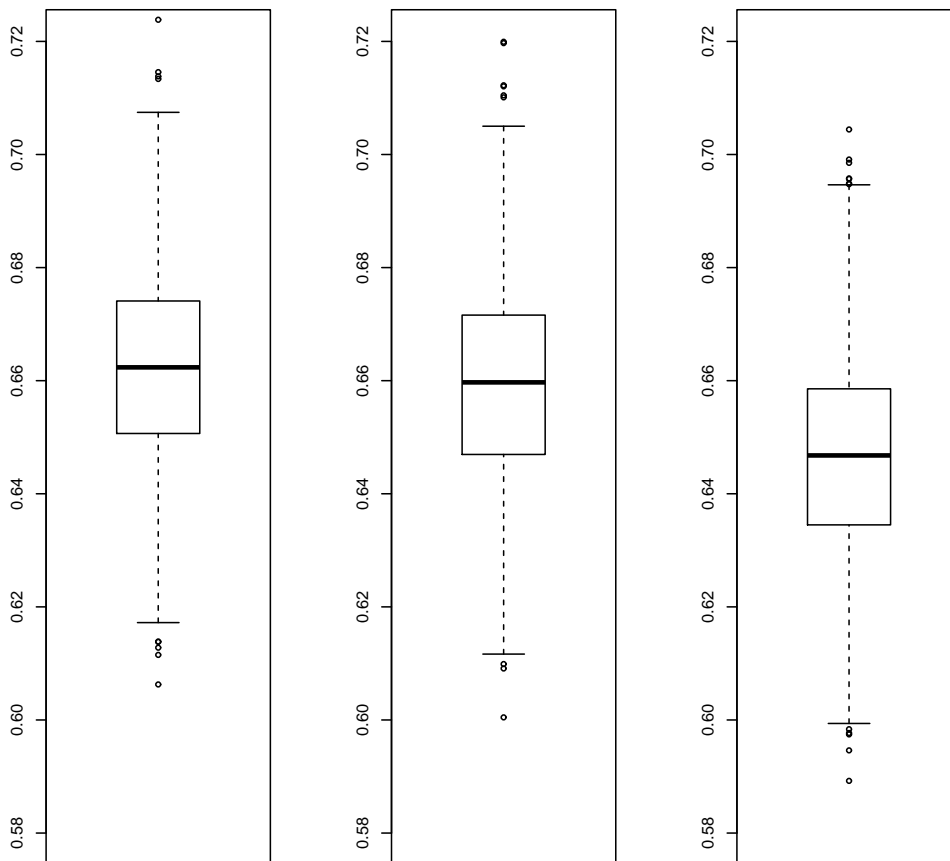


Figure 6: The bootstrap distribution of the *AUC* corresponding to 25 % embryo data for validation for EQI, LMCP and Steer

In a similar way to the above subsection we calculated confidence intervals

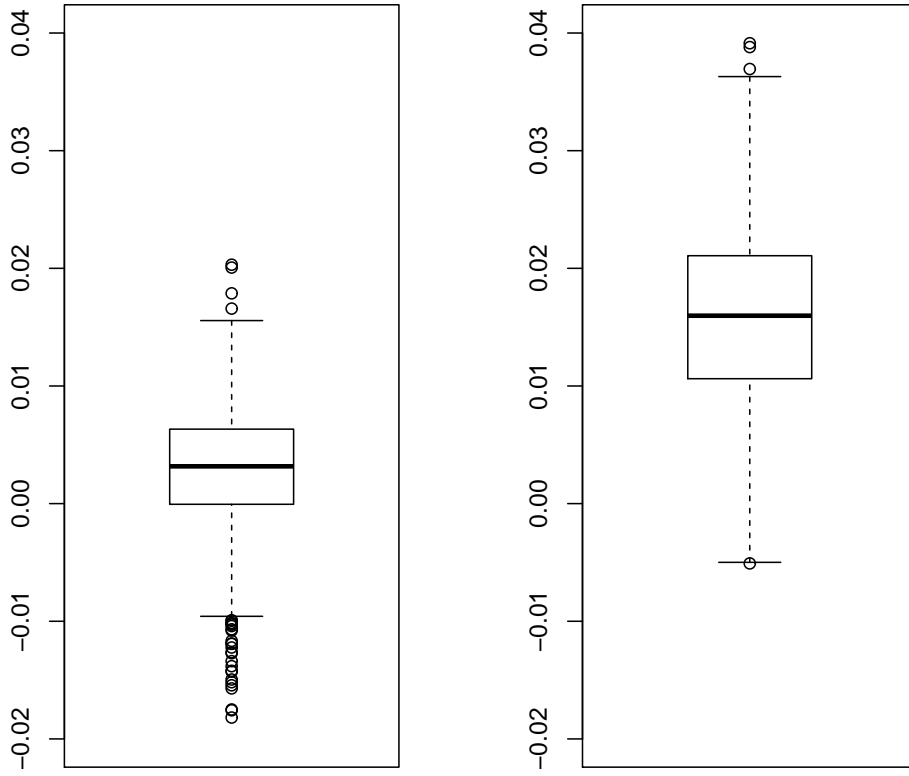


Figure 7: The bootstrap distribution of the difference of the AUCs corresponding to 25 % embryo data for validation for EQI vs LMCP and EQI vs CES

(CI) for AUC s and their differences, and we applied a paired t-test evaluated with EQI as the model A and each of LMCP and CES as model B. Again the

application of paired t-test was supported by the normality of the difference between the AUCs. Table 9 shows the results. According to these t-values and their corresponding p-values it can be concluded that the “true” mean of *AUCs* of EQI are greater than the “true” mean of *AUCs* of the LMCP and CES. Therefore, using any of the logistic regression models should outperform a simpler scoring system such as Steer *et al.*[11] and treating the variables: number of blastomeres and grade as categorical variables means a slightly more discriminatory capacity for the model.

Model	CI(95%)	CI(95%)	t-value	p-value
EQI	[0.6269,0.6964]			
CES	[0.6115,0.6805]	[0.0002,0.0308]	124.04	0.0027
LMCP	[0.6227,0.6930]	[-0.0096,0.0120]	52.09	0.0159

Table 9: The confidence intervals (CI) for AUC’s and their differences and the results of t-test corresponding to 25 % embryo data for validation

3.4. Impact of women’s age and EQI embryo scoring model on their pregnancy

The effects of EQI and women’s age on the implantation rate, calculated as the number of sacs per transferred embryos, were analyzed using GLM with

continuous predictors. The corresponding implantation rate for each woman was treated as a binomial variable although in this study we obtained a triple pregnancy by transfer of only two embryos, it can be considered as “both implantations were successful”. For these reasons, the Binomial distribution with link logit is used again in a GLM. To obtain the intercept of this model we transformed the variables by subtracting the minimum of their values, thus $tage = age - 20$ and $tEQI = EQI - 0.32$. This model was fitted using the same `glm` function but the R code was:

```
glm(cbind(sacos,Embryo.transfer-sacos)+tage+tEQI+(tage:tEQI),
family=binomial(link=log), data=women),
```

where `sacos` and `Embryo.transfer` are the number of gestational sacs, the number of transferred embryos and the database is `women`. Although the saturated model (main effects plus interaction) was fitted, the model with only main effects and without interaction was selected by using `step` command. These results are shown in Table 10.

The intercept is the logit of the implantation rate corresponding to a 20 year old woman with a minimum embryo score of 0.34. The value -0.0553 measures the decrease implied in the logit implantation rate logarithm when a woman’s age increases by a year, which is when the score remains con-

stant. 0.0402 is the increase implied in this logarithm when the embryo score increases by one unit when age remains constant. Then, we can check model assumptions. The first standard graphs are shown in Figure 8 using R command `plot`. Figure 8 shows the plot of the deviance residuals versus the predicted values in the upper left panel, the dispersion of residuals fluctuates around zero but the variance does not remain more or less constant with the average. In the top right panel there is a Q-Q plot of standardized deviance residuals. If these residuals are adjusted perfectly to the diagonal, their distribution is exactly normal. Our residuals are separated from the line, especially at the ends, which means that the distribution of these residuals has thicker tails than the Normal distribution. The graph in the lower left panel is the representation of the square root of the absolute value of the deviance residuals versus predicted values. Curvatures will indicate the absence of a quadratic term or a bad choice of link function. In this case, it seems that the data does not reasonably fit a straight line as there are some points that are more spread for that reason we need to check the linear relationship using the GAM model. The last graph, located in the lower right panel, determines the influential points using Cook's distances, which is a measure of the difference between the fitted values of the model and the

model without each of the observations.

Parameter	Estimate	SE	t-value	p-value
Intercept	-1.5742	0.1212	-13.00	0
tAge	-0.0553	0.0073	-7.53	0
tEQI	0.0402	0.0024	16.52	0

Table 10: Estimates for the parameters of GLM with main effects of continuous predictors: tAge and tEQI.

In order to check the consistency of these previous results we used GAM model, technique which allows us to incorporate explanatory variables in a non-linear way in the model. The resulting plot of the GAM model is shown in Figure 9 where smooth functions are written as $s(x, df)$, x is the univariate predictor, and df is the target equivalent degrees of freedom, used as a smoothing parameter (values for df should be greater than 1, with $df = 1$ implying a linear fit). Figure 9 suggests a non-linear relationship for both variables, therefore indicating that the linearity that the GLM model supposes is not confirmed.

Normally the predicted implantation rate is obtained by using logistic regression [18] but results from GLM are strongly dependent on the linear

assumption. For example, a woman to whom a 4-blastomere grade 1 embryo has been transferred, whose score is 13.47 at age 20, has predicted implantation rate of 0.2600 with GLM but 0.1851 with GAM, while at age 40 this implantation rate is only 0.1042 with GLM but 0.0904 with GAM. In the case of a maximum score, that is to say that the woman has a transfer of three embryos of 4-blastomere grade 1, whose total score is 40.41, the predicted implantation rate at age 20 is 0.5097 with GLM but 0.3473 with GAM, whereas at age 40 it is only 0.2559 with GLM but 0.1890 with GAM.

4. Discussion

Although there is general agreement among embryologists as to which morphological features are characteristic of a “top” embryo in the cleavage stage, evidence is still lacking for the ranking of implantation potential of non-top embryos. The need to establish greater knowledge about embryo quality variables and thus construct reliable scoring systems is becoming increasingly evident.

The reason for the lacking of scientific data is largely due to the difficulties in following the fate of an individual embryo. The prevailing clinical practice of transferring more than one embryo makes deduction from embryo quality

variables unreliable when the resulting pregnancy contains fewer sacs than the number of transferred embryos. Therefore, the available scientific data to date is based on studies containing a limited number of treatments with a traceable association between embryo and implantation [13, 14, 18, 19, 44].

The ideal approach to studying the morphological determinants of a single embryo's implantation potential would be to exclusively analyse single-embryo transfers. However, in most single-embryo transfer programmes only "top" embryos are transferred, and thus an optimal span in variables for statistical evaluation cannot be obtained by this approach. Alternatively, data from treatments, which result in only a single embryo available for transfer should also be analysed. Although, this has been carried out, producing important information, the evaluation of such data is hampered by the fact that these treatments mainly involve women with a poor response, poor embryo quality and low implantation figures, thus again not allowing a wide span of morphological variation [13,18]. The strategy of our study was similar to those applied in the retrospective analyses published by Ziebe et al. [44] (cycles in which the two or three transferred embryos were of identical morphological score) and by Van Royen et al. [14, 19] (all pregnancies of equal implantation number as the number of embryos transferred) and those

of Holte et al [18], who performed a prospective study on a large number of embryos with exact fate, which made it possible to produce very powerful statistics. In this study we used all the cycles with transfers of one, two and three embryos with an exact fate (0, 1, 2 or 3 ultrasonically identified gestational sacs) to construct the model, validating it in a second step by including transfers with a lower number of ultrasonically identified sacs than transferred embryos. As we only included two morphological embryo variables (blastomere number and grade), we found a relatively simple model with two real-time visual embryo-scoring variables, but the implantation potential of embryos is also related to various clinical characteristics of the women in addition to cleavage stage embryo morphology as discussed in the paper. Therefore, the present study does not propose an implantation prediction system but an embryo scoring model. We confirm that our model ranks the ability to achieve pregnancy better than the CES proposed by Steer et al. [11]. In relation to other methodologies such as LMCP, the methodology employed here has allowed us to identify differences between each value for the number of blastomeres and each grade value.

Cleavage rate has turned out to be a powerful marker for implantation potential, corroborating other reports on this variable's importance [13, 14,

18, 19, 44]. Many authors have used the number of embryos transferred and their CES in the prediction of pregnancy [11, 27, 29, 30, 45, 46].

With regard to the study by Laasch and Puscheck [29], they conclude that if the CES is under 50, the transfer of one more embryo can be considered in order to increase the additional CES, showing that the score for each embryo is a key issue. In this study, we have improved the score for each embryo and the threshold score from which we can distinguish between pregnant women and those not pregnant, so we can better adjust the number of embryos transferred and thereby avoid possible multiple pregnancies.

In our study, the same as in previous studies, we used embryo scoring systems that did not permit the scoring of both blastomere size variation and fragmentation in a single embryo, because the system, through its rigidity, forces the embryologist to score for either one or the other variable, but not both simultaneously. In general, grade 1 and 2 embryos had almost equally-sized blastomeres with low fragmentation, while grade 3 and 4 embryos had unequally sized blastomeres with moderate and severe fragmentation, and only these last embryos had a significantly lower implantation rate. In relation to the blastomere fragmentation, Holte et al [18] and Saldeen and Sundstrom [47] did not observe any significant differences for the implanta-

tion rate. In this last study the authors evaluated the blastomere fragmentation as an independent embryo variable and fragmentation was insignificant. Saldeen and Sundstrom [47] obtained the same results, but in that study the material consisted exclusively of 4-cell embryos in elective single-embryo transfers, and thus embryos of a generally higher quality and lower morphological variation than in the present study.

In any case, our results agree with those obtained by several authors who have previously observed that slight fragmentation does not have a negative impact on implantation [13, 14, 44, 48], and small fragments may disappear through lysis or resorption during culture [15, 49]. Fragmentation was not associated with a chromosomal abnormality rate in a recent study [48]. A superiority of the cleavage stage over the grade of fragmentation for judging embryo competence has also been suggested by several researchers [13, 14, 44, 48].

It is possible that the scoring of fragmentation should also take into account the localization of fragments [22, 49, 50] and that the occurrence of fragmentation at later cleavage stages has more biological relevance [23, 50], and only severe fragmentation has been associated with the increased occurrence of malformations [22].

It should be emphasised that all visual real-time scoring procedures are affected by varying inherent difficulties i.e. the intra- and inter-observer variations are likely to be larger for some variables than others. Such qualities in a parameter may diminish its prognostic power, even if the variable is of significant biological importance. The grade variable is presumably the least precise variable in this sense. At the other extreme is the cleavage rate variable. The low risk of misscoring the cleavage rate is probably one of the factors that put this variable into the best scoring position. Therefore, it should be borne in mind that the scoring model derived from the present study should be regarded primarily as a clinical tool for selecting embryos to transfer. The biological significance of the findings must be supported and tested in further prospective studies. The validation of our proposed model was performed using the ROC curve that provides a comprehensive representation of the accuracy of the method, and not by the comparison of the means of the groups such as Holte et al. [18].

In our work we have included age and the measurement of embryo quality as continuous variables in a GLM and to check the relationship in a GAM. Holte et al. [18] conducted a similar analysis, though we must emphasize that although those authors chose grouped (with 6 groups) age variable which is

included only in a linear way in the model. In accordance with these authors age does not interact with any of the variables i.e. the model can be applied to any woman independent of age, but here the implantation figures decrease non-proportionally with age.

However, for the purpose of a full clinical prediction model for implantability, the embryo scoring model obtained by the present study (EQI) must be combined with other significant demonstrated variables for the implantation rate such as the measurement of ovarian sensitivity to FSH reflecting ovarian age, the estradiol levels in the HCG day reflecting the ovarian response to the controlled ovarian hyperstimulation protocol (COH), the number of available embryos reflecting the fertilization rate efficiency and the intensity of embryo selection prior to transfer; since the higher the number of embryos obtained, the greater the possibility of transferring embryos of better quality. In conclusion, our model provides a tool that allows quick decisions to be made when choosing the best embryos for transfer. We have also proposed the ROC curve as a graphical tool and the AUC as a numerical value for validation and comparison of the different models. In addition we proposed a test which compared our model with other previous models demonstrating its superiority. This model can also be used in those databases in which, like

ours, equality and symmetry variables are grouped into the grade variable.

ACKNOWLEDGEMENTS

This work was partially supported by a grant from the Generalitat Valenciana (grant No. GVPRE/2008/103). The research of Ana Debón and Suitberto Cabrera was partially supported by a grant from Ministerio de Asuntos Exteriores (grant No. A/023444/09) too. The authors are indebted to the anonymous referee whose comments suggestions improved the paper considerably.

- [1] T. Bergh, A. Ericson, T. Hillensjo, K. Nygren, U. Wennerholm, Deliveries and children born after in vitro fertilisation in Sweden 1982-95: a retrospective cohort study, *Lancet* 354 (4) (1999) 1579–1585.
- [2] B. Stromberg, G. Dahlquist, A. Ericson, O. Finnstrom, M. Koster, K. Stjernqvist, Neurological sequelae in children born after in vitro fertilisation: a population-based study, *Lancet* 359 (9305) (2002) 461–465.
- [3] V. Wright, J. Chang, G. Jeng, M. Macaluno, Assisted reproductive technology surveillance-United States 2003, *MMWR Surveillance Summaries* 55 (SS-4) (2006) 1–22.

- [4] Y. Hu, W. Maxson, D. Hoffman, Maximizing pregnancy rates and limiting higher-order multiple conceptions by determining the optimal number of embryos to transfer based on quality, *Fertility and Sterility* 69 (4) (1998) 650–657.
- [5] A. Strandell, C. Bergh, K. Lundin, Selection of patients suitable for one-embryo transfer reduces the rate of multiple births by half without impairment of overall birth rates, *Human Reproduction* 15 (12) (2000) 2520–2525.
- [6] J. Catt, T. Wood, M. Henman, R. Jansen, Single embryo transfer in IVF to prevent multiple pregnancies, *Twin Research* 6 (6) (2003) 536–539.
- [7] D. D. Neubourg, J. Gerris, K. Mangelschots, E. V. Royen, M. Ver-cruyssen, M. Elseviers, Single top quality embryo transfer as a model for prediction of early pregnancy outcome, *Human Reproduction* 19 (6) (2004) 1476–1479.
- [8] A. Van Montfoort, J. Dumoulin, J. Land, E. Coonen, J. Derhaag, J. Evers, Elective single embryo transfer (eSET) policy in the first three ivf/icsi cycles treatment cycles, *Human Reproduction* 20 (2) (2005) 433–436.

- [9] C. Bergh, Single embryo transfer: a mini-review, *Human Reproduction* 20 (2) (2005) 323–327.
- [10] D. D. Neubourg, J. Gerris, What about the remaining twins since single-embryo transfer? how far can (should) we go?, *Human Reproduction* 21 (4) (2006) 843–846.
- [11] C. Steer, C. Mills, S. Tan, The cumulative embryo score: a predictive embryo scoring technique to select the optimal number of embryos to transfer in a in vitro fertilization and embryo transfer programme, *Human Reproduction* 7 (1) (1992) 117–119.
- [12] A. Shulman, I. Ben-Nun, Y. Ghetler, H. Kaneti, M. Shilon, Y. Beyth, Relationship between embryo morphology and implantation rate after in vitro fertilization treatment in conception cycles, *Fertility and Sterility* 60 (1) (1993) 123–126.
- [13] C. Giorgetti, P. Terriou, P. Auquier, E. Hans, J. Spach, J. Salzmann, R. Roulier, Embryo score to predict implantation after in vitro fertilization: based on 957 single embryo transfers, *Human Reproduction* 10 (9) (1995) 2427–2431.
- [14] E. Van Royen, K. Mangelschots, D. D. Neubourg, M. Valkenburg, M. V.

- de Meerssche, G. Ryckaert, W. Eestermans, J. Gerris, Characterization of a top quality embryo, a step towards single-embryo transfer, *Human Reproduction* 14 (9) (1999) 2345–2349.
- [15] T.Hardarson, C. Hanson, A. Sjogren, K. Lundin, Human embryos with unevenly sized blastomeres have lower pregnancy and implantation rates: indications for aneuploidy and multinucleation, *Human Reproduction* 16 (2) (2001) 313–318.
- [16] P. Terriou, C. Sapin, C. Giorgetti, E. Hans, J. Spach, R. Roulier, Embryo score is a better predictor of pregnancy than the number of transferred embryos or female age, *Fertility and Sterility* 75 (3) (2001) 525–531.
- [17] C. Hunault, M. Eijkemans, M. Pieters, E. te Velde, J. Habbema, B. Fauser, N. Macklon, A prediction model for selecting patients undergoing in vitro fertilization for elective single embryo transfer, *Fertility and Sterility* 77 (4) (2002) 725–732.
- [18] J. Holte, L. Berglund, K. Milton, C. Garello, G. Gennarelli, A. R. A, T. Bergh, Construction of an evidence-based integrated morphology cleavage embryo score for implantation potential of embryos scored

and transferred on day 2 after oocyte retrieval, *Human Reproduction* 22 (2) (2007) 548–557.

- [19] E. Van Royen, K. Mangelschots, D. De Neubourg, I. Laureys, G. Ryckaert, J. Gerris, Calculating the implantation potential of day 3 embryos in women younger than 38 years of age: a new model, *Human Reproduction* 16 (2) (2001) 326–332.
- [20] E. V. Royen, K. Mangelschots, M. Vercruyssen, D. D. Neubourg, M. Valkenburg, G. Ryckaert, J. Gerris, Multinucleation in cleavage stage embryos, *Human Reproduction* 18 (5) (2003) 1062–1069.
- [21] L. Scott, The biological basis of non-invasive strategies for selection of human oocytes and embryos, *Human Reproduction* 9 (3) (2003) 237–249.
- [22] T. Ebner, C. Yaman, M. Moser, M. Sommergruber, W. Polz, G. Tews, Embryo fragmentation in vitro and its impact on treatment and pregnancy outcome, *Fertility and Sterility* 76 (2) (2001) 281–285.
- [23] T. Ebner, M. Moser, M. Sommergruber, G. Tews, Selection based on morphological assessment of oocytes and embryos at different stages of

- preimplantation development: a review, *Human Reproduction Update* 9 (3) (2003) 251–262.
- [24] L. Rienzi, F. Ubaldi, M. Iacobelli, S. Romano, M. Minasi, S. Ferrero, F. Sapienza, E. Baroni, E. Greco, Significance of morphological attributes of the early embryo, *Reproductive biomedicine online* 10 (5) (2005) 669–681.
- [25] L. Scott, A. Finn, T. O’Leary, S. McLellan, J. Hill, Morphologic parameters of early cleavage-stage embryos that correlate with fetal development and delivery: prospective and applied data for increased pregnancy rates, *Human Reproduction* 22 (1) (2007) 230–240.
- [26] T. Child, A. Abdul-Jalil, S. Tan, Embryo morphology, cumulative embryo score, and outcome in an oocyte in vitro maturation program, *Fertility and Sterility* 77 (2) (2002) 424–425.
- [27] A. Copperman, C. Selick, L. Grunfeld, B. Sandler, M. Bustillo, Cumulative number and morphological score of embryos resulting in success: realistic expectations from in vitro fertilization-embryo transfer, *Fertility and Sterility* 64 (1) (1995) 88–92.
- [28] J. Cummins, T. Breen, K. Harrison, J. Shaw, L. Wilson, J. Hennessey, A

formula for scoring human embryo growth rates in vitro fertilization: its value in predicting pregnancy and in comparison with visual estimates of embryo quality, *Journal of In vitro Fertilization and Embryo Transfer*. 1986 3 (5) (1986) 284–295.

[29] C. Laasch, P. E. J. Cummins, T. Breen, K. Harrison, J. Shaw, L. Wilson, J. Hennessey, Cumulative embryo score, not endometrial thickness, is best for pregnancy prediction in IVF, *Journal of Assisted Reproduction and Genetics* 21 (2) (2004) 47–50.

[30] D. Visser, F. Fourie, The applicability of the cumulative embryo score system for embryo selection and quality control in an in vitro fertilization/embryo transfer programme, *Human Reproduction* 8 (10) (1993) 1719–1722.

[31] I. Molina, J. Alfonso, C. Duque, L. García-Reboll, M. Pérez-Camps, A. Romeu, Severe oligoasthenoteratozoospermias, secretory and obstructive azoospermias: Motility as a criterion of sperm viability, *Journal of Assisted Reproduction and Genetics* 24 (4) (2006) 125–129.

[32] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna,

Austria, ISBN 3-900051-07-0 (2009).

URL <http://www.R-project.org>

- [33] T. Sing, O. Sander, N. Beerenwinkel, T. Lengauer, ROCR: Visualizing the performance of scoring classifiers., r package version 1.0-2 (2007).

URL <http://rocr.bioinf.mpi-sb.mpg.de/>

- [34] P. McCullagh, J. Nelder, Generalized Linear Models, Chapman and Hall, London, 1989.

- [35] A. Agresti, Categorical Data Analysis, Wiley, 2^a Edición, 2002.

- [36] A. F. Zuur, E. N. Ieno, N. J. Walker, A. A. Saveliev, G. M. Smith, Mixed effects models and extensions in ecology with R, Springer, 2009.

- [37] D. Bates, M. Maechler, B. Bolker, lme4: Linear mixed-effects models using S4 classes, r package version 0.999999-0 (2012).

URL <http://CRAN.R-project.org/package=lme4>

- [38] T. Fawcett, An introduction to ROC analysis, Pattern Recognition Letters 27 (8) (2006) 861–874.

- [39] B. Efron, R. Tibshirani, An introduction to the bootstrap, Chapman & Hall, New York & London, 1993.

- [40] Z. Zhang, M. Pepe, A linear regression framework for receiver operating characteristic(roc) curve analysis, Tech. rep., Join COBRA (2005).
- [41] J. A. Hanley, B. J. McNeil, A method of comparing the areas under receiver operating characteristic curves derived from the same cases. 148 (3) (1983) 839–843.
- [42] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning. Data Mining, Inference, and Prediction, Springer, New York, 2001.
- [43] J. de Leeuw, P. Mair, Simple and canonical correspondence analysis using the R package anacor, Journal of Statistical Software 31 (5) (2009) 1–18.
URL <http://www.jstatsoft.org/v31/i05/>
- [44] S. Ziebe, K. Petersen, S. Lindenberg, A. G. Andersen, A. Gabrielsen, A. N. Andersen, Embryo morphology or cleavage stage: how to select the best embryos for transfer after in-vitro fertilization., Human Reproduction 12 (7) (1997) 1545–1549.
- [45] N. L. Dean, S. J. Phillips, W. M. Buckett, M. M. Biljan, S. L. Tan, Impact of reducing the number of embryos transferred from three to

two in women under the age of 35 who produced three or more high-quality embryos, *Fertility and Sterility* 74 (4) (2000) 820 – 823.

- [46] J. D. Fisch, H. Rodriguez, R. Ross, G. Overby, G. Sher, The Graduated Embryo Score (GES) predicts blastocyst formation and pregnancy rate from cleavage-stage embryos, *Human Reproduction* 16 (9) (2001) 1970–1975.
- [47] S. P. S. P., Nuclear status of four-cell preembryos predicts implantation potential in in vitro fertilization treatment cycles, *Fertility and Sterility* 84 (2005) 584–589.
- [48] S. Ziebe, K. Lundin, A. Loft, C. Bergh, A. Nyboe Andersen, U. Sellskog, D. Nielsen, C. GrÃ¸ndahl, H. Kim, J. Arce, Fish analysis for chromosomes 13, 16, 18, 21, 22, x and y in all blastomeres of ivf pre-embryos from 144 randomly selected donated human oocytes and impact on pre-embryo morphology, *Human Reproduction* 18 (12) (2003) 2575–2581.
- [49] J. V. Blerkom, P. Davis, S. Alexander, A microscopic and biochemical study of fragmentation phenotypes in stage-appropriate human embryos, *Human Reproduction* 16 (4) (2001) 719–729.

- [50] M. Alikani, J. Cohen, G. Tomkin, G. J. Garrisi, C. Mack, R. T. Scott,
Human embryo fragmentation in vitro and its implications for pregnancy
and implantation, *Fertility and Sterility* 71 (5) (1999) 836 – 842.

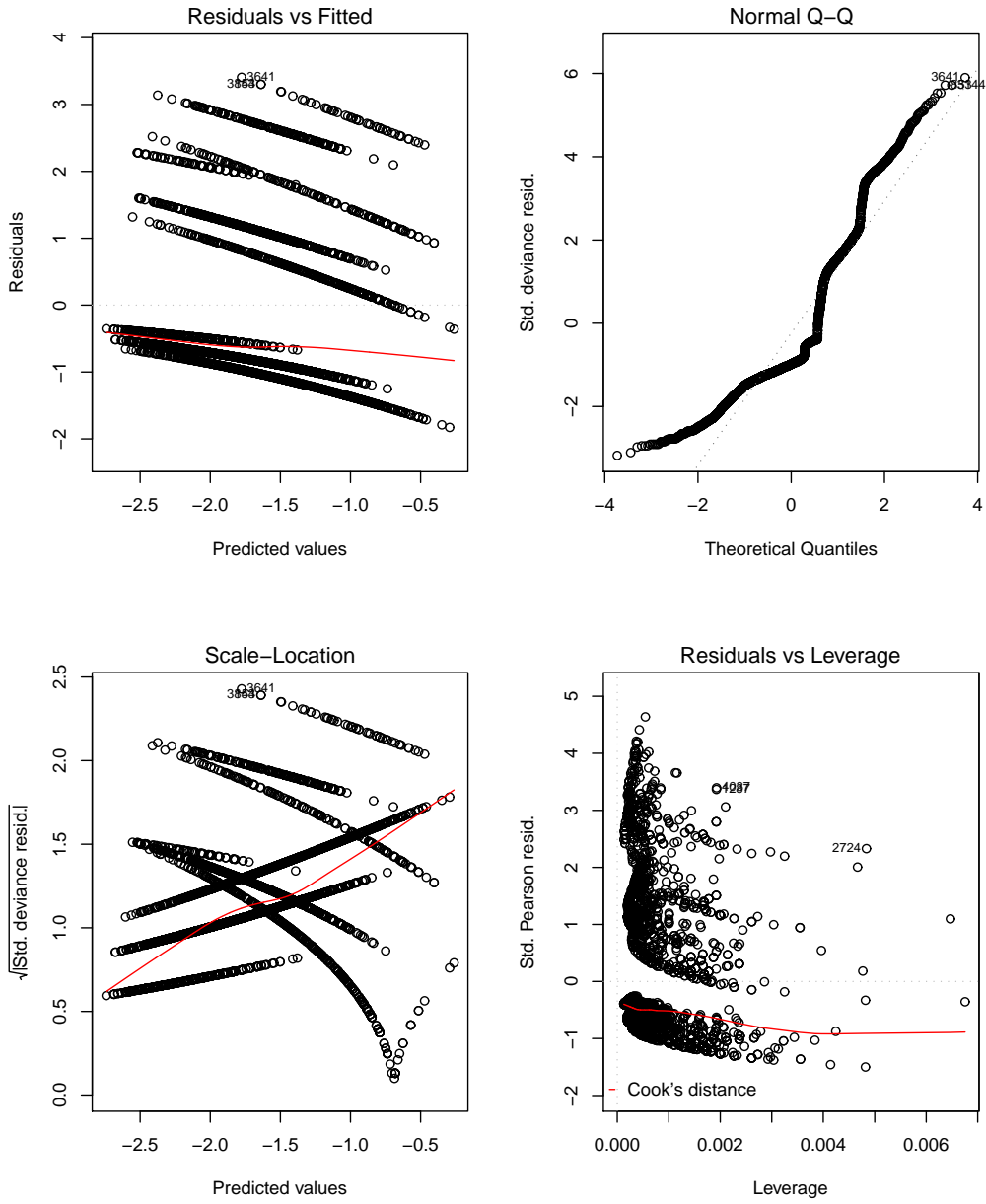


Figure 8: Checking the assumptions of women's age and EQI for the GLM model.

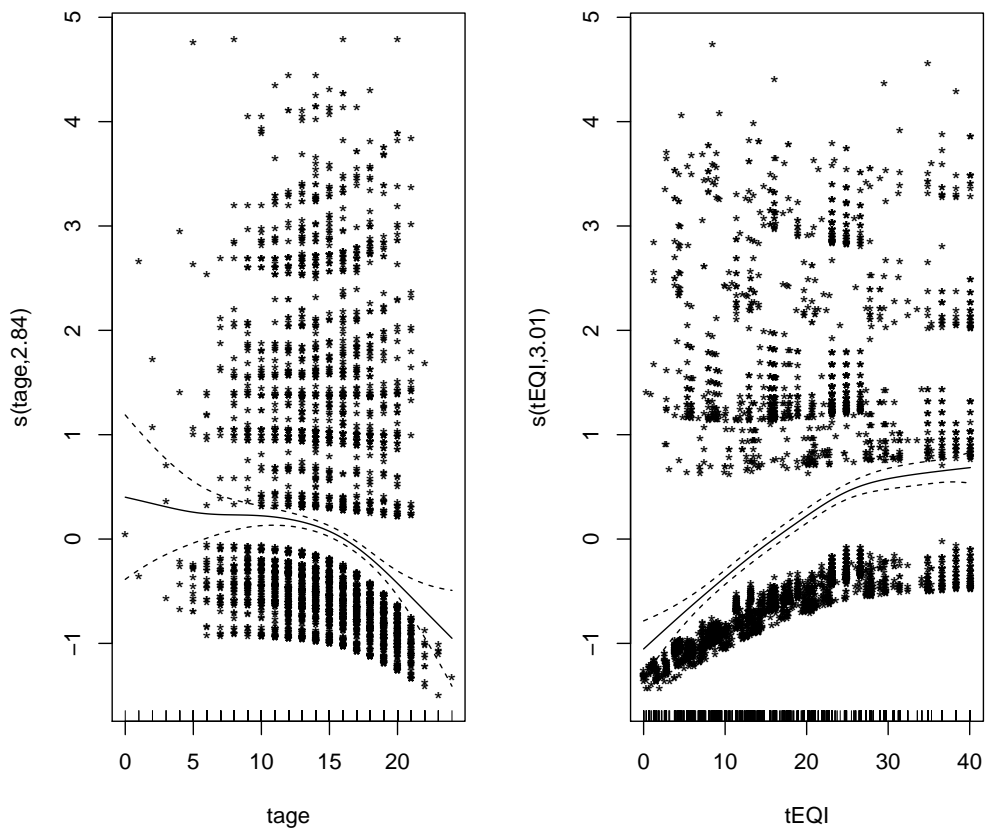


Figure 9: Plots of the component smooth functions and partial residuals on the scale of the linear predictor