Reproductive BioMedicine Online (2011) **22**, 341–349

www.sciencedirect.com
www.rbmonline.com

ARTICLE

Anti-Müllerian hormone-based prediction model for a live birth in assisted reproduction

A La Marca [a,*], SM Nelson [c], G Sighinolfi [a], M Manno [d], E Baraldi [a], L Roli [a], S Xella [a], T Marsella [a], D Tagliasacchi [a], R D'Amico [b], A Volpe [a]

[a] *Mother–Infant Department, Institute of Obstetrics and Gynecology, University of Modena and Reggio Emilia, 41100 Modena, Italy;* [b] *Department of Oncology, Hematology and Respiratory Diseases, University of Modena and Reggio Emilia, 41100 Modena, Italy;* [c] *Reproductive and Maternal Medicine, Faculty of Medicine, University of Glasgow, Glasgow, UK;* [d] *Maternal-Paediatric Department, Pathophysiology of Human Reproduction and Sperm Bank Unit, Pordenone Hospital, Pordenone, Italy*

* Corresponding author. *E-mail address:* antlamarca@libero.it (A La Marca).

Dr Antonio La Marca graduated in medicine in 1996 and completed his residency in obstetrics and gynaecology in 2001 and his PhD on biology of germinal cells in 2004 at University of Siena, Italy. Since then, he has worked at the Mother–Infant Department, University Hospital of Modena, Italy. His research areas are the physiology of ovarian function and its pharmacological manipulation. In recent years his research has been focused on anti-Müllerian hormone and ovarian reserve.

Abstract Prediction of assisted reproduction treatment outcome has been the focus of clinical research for many years, with a variety of prognostic models describing the probability of an ongoing pregnancy or a live birth. This study assessed whether serum anti-Müllerian hormone (AMH) concentrations may be incorporated into a model to enhance the prediction of a live birth in women undergoing their first IVF cycle, by analysing a database containing clinical and laboratory information on IVF cycles carried out between 2005 and 2008 at the Mother–Infant Department of University Hospital, Modena. Logistic regression was used to examine the association of live birth with baseline patient characteristics. Only AMH and age were demonstrated in regression analysis to predict live birth, so a model solely based on these two criteria was generated. The model permitted the identification of live birth with a sensitivity of 79.2% and a specificity of only 44.2%. In the prediction of a live birth following IVF, a distinction, however moderate, can be made between couples with a good and a poor prognosis. The success of IVF was found to mainly depend on maternal age and serum AMH concentrations, one of the most relevant and valuable markers of ovarian reserve.

KEYWORDS: AMH, assisted reproduction, live birth, ovarian reserve, prediction model

Introduction

The prediction of assisted reproduction treatment outcome has been the focus of clinical research for many years, with a variety of prognostic models describing the probability of an ongoing pregnancy or a live birth following treatment. To date, models have predominantly been established using patient baseline characteristics and although these models

1472-6483/$ - see front matter
doi:10.1016/j.rbmo.2010.11.005

have been heterogeneous in their performance they consistently demonstrate that certain patient characteristics are associated with IVF/intracytoplasmic sperm injection (ICSI) success, including female age (Bancsi et al., 2000; Bouckaert et al., 1994; Carrera-Rotllan et al., 2007; Commenges-Ducos et al., 1998; Hughes et al., 1998; Hunault et al., 2002; Hull et al., 1996; Lee et al., 2009; Lintsen et al., 2005; Minaretzis et al., 1998; Ottosen et al., 2007; Stolwijk et al., 1996; Stolwijk et al., 2000; Templeton et al., 1996; van Weert et al., 2008; Younis et al., 2009), duration of infertility (Haan et al., 1991; Lintsen et al., 2007; Younis et al., 2009), pregnancy history (Lintsen et al., 2007; Stolwijk et al., 2000; Templeton et al., 1996; van Weert et al., 2008), diagnostic category (Bancsi et al., 2000; Lintsen et al., 2007; van Weert et al., 2008) and body mass index (BMI; Ferlitsch et al., 2004; Verberg et al., 2008).

Alternative models have incorporated the characteristics of the intermediate results of the first treatment cycle thereby improving the accuracy of probability estimates for future cycles. The variables used in these models include the number of retrieved oocytes (Bouckaert et al., 1994; Hunault et al., 2002; Verberg et al., 2007), the fertilization rate and embryo number and quality (Hunault et al., 2002; Minaretzis et al., 1998; Ottosen et al., 2007; Verberg et al., 2007). Moreover it has been clearly demonstrated that in models predicting pregnancy based on intermediate results (at embryo transfer), the number of retrieved and fertilized oocytes have the highest prognostic value (Ferlitsch et al., 2004; Verberg et al., 2007). This suggests that any marker which can predict the number of retrieved oocytes prior to ovarian stimulation may be of value in initial baseline prognostic models (Bancsi et al., 2000; Carrera-Rotllan et al., 2007; Ferlitsch et al., 2004; Lee et al., 2009; Ottosen et al., 2007; Younis et al., 2009).

Recent studies have indicated that anti-Müllerian hormone (AMH) may constitute an important novel measure of ovarian reserve, with the current literature indicating that AMH is a superior marker for predicting ovarian response over either age of the patient, day-3 FSH, oestradiol or inhibin B levels, whereas the vast majority of studies have found AMH and antral follicle count to have similar predictive value (for review, see La Marca et al., 2009). Consistent with AMH being a strong correlate of oocyte yield, AMH has recently been proposed as a useful clinical marker for the prediction of both poor- and hyperresponses to ovarian stimulation (La Marca et al., 2009). In addition to reflecting the quantitative ovarian response, several authors have found a significant positive correlation between AMH concentrations and oocyte quality (Cupisti et al., 2007; Ebner et al., 2006; Hazout et al., 2004; Silberstein et al., 2006), fertilization rate (Lekamge, 2007) and embryo morphology (Silberstein et al., 2006). However, this relationship has not been confirmed by others (Lie Fong et al., 2008; Smeenk et al., 2007). Hence the possible prediction of qualitative aspects of assisted reproduction programmes by AMH measurement is largely controversial.

In one large study, AMH was shown to be associated with live birth independent of age after treatment (Nelson et al., 2007). More recently a further large cohort study demonstrated that serum AMH concentrations may predict live birth in women older than 34 (Lee et al., 2009). The aim of the present study was to assess whether serum AMH concentrations may be incorporated into a prediction model to enhance the prediction of a live birth in women undergoing their first IVF. In particular, the objective was to develop a simple multivariate score based on basal patients characteristics which was capable of predicting the outcome of the treatment cycle and to express this in a clean format which could be easily adopted into daily clinical practice.

Materials and methods

Study population

This study analysed the database containing clinical and laboratory information on IVF treatment cycles carried out at the Mother–Infant Department of University Hospital, Modena between 2005 and 2008. These data were collected prospectively and recorded in the registered database in the fertility centre in Modena, Italy. Cycles were selected for analysis if all the following inclusion criteria were satisfied: (i) first IVF/ICSI attempt; (ii) a normal uterus and regular uterine cavity; (iii) no previous ovarian surgery; (iv) absence of severe male factor (defined as sperm count less than 1×10^6/ml or normal forms less than 5% according to World Health Organization (1999); (v) female age $\leq$42; (vi) absence of recurrent abortion; (vii) absence of antiphospholipid syndrome and any other relevant systemic condition; (viii) treatment with a long gonadotrophin-releasing hormone (GnRH) agonist protocol; (ix) complete computer-based patient records on anamnestic, clinical and IVF cycle characteristics and pregnancy follow-up; and (x) a stored serum sample taken within 3 months of commencing IVF suitable for measurement of AMH. All patients had been trying to conceive for at least 12 months and all had undergone a fertility workup.

The long GnRH agonist protocol (Enantone; Takeda Italia, Rome, Italy) is based on the administration of leuprorelin on day 21 of the previous luteal phase of the stimulation cycle. Recombinant FSH at a dose ranging between 150 and 300 IU/day subcutaneously was commenced on cycle days 2–3 and then the dose was adjusted on days 7–8 according to the ovarian response. When at least two follicles reached $>$18 mm, 10,000 IU of human chorionic gonadotrophin was administrated intramuscularly and 34–36 h later follicles were aspirated under patient sedation. Insemination was performed by standard IVF or ICSI. According to the new Italian law regulating assisted reproduction treatment, only three oocytes were fertilized at one time. Light microscopic evaluation established fertilization 14–18 h later. Cleavage-stage embryo transfers were performed on day 2 or 3 under ultrasound guidance. A serum human chorionic gonadotrophin pregnancy test was performed 14 days after retrieval and repeated 7 days later if positive.

Clinical pregnancy was defined as ultrasound visualization of a gestational sac with evidence of a fetal heart and excluded all ectopic and biochemical pregnancies. Live birth was defined by the birth of at least one live-born child.

All patients gave written informed consent at the time of the IVF cycle for both the procedure and for digital recording and the use of laboratory and clinical data related to their medical history for clinical research purposes.

AMH assay

Blood samples were taken between 8:00 a.m. and 12:00 a.m. from the cubital vein, in the early follicular phase prior to any IVF-related drug administration. The blood was centrifuged at 2000*g* for 10 min and the serum was stored in polypropylene tubes at −80°C. Serum AMH was measured by enzyme-linked immunosorbent assay (ELISA) using the Beckman Coulter AMH ELISA kit (Immunotech, Marseilles, France). The detection limit of the assay was 0.14 ng/ml; intra- and inter-assay coefficients of variation were 12.3% and 14.2%, respectively (conversion factor: 1 ng/ml = 7.14 pmol/l). The immunoassay is specific for AMH. No cross-reaction was observed with transforming growth factor β.

Statistical analysis

The endpoint of the study was to identify factors associated with live birth and develop a clear model which could be easily adopted into daily clinical practice. To date, the majority of prediction models in IVF are represented by complicated formulae which cannot be used by clinicians. To facilitate the development of a useful model, although initial exploratory analysis used all variables as continuous predictors, age and AMH were subsequently treated as categorical variables. The age groups were stratified based upon both the physiological understanding of natural fertility, whose initial decline begins at 31 years (te Velde and Pearson, 2002) and the critical age of 37 years recorded as the pivotal age for success rates in treatment programmes (Human Fertilisation and Embryology Authority/Society for Assisted Reproductive Technology). Since its distribution was found not to be normal, AMH was log transformed. AMH was then stratified into three classes according to 25th and 75th centile for its values in the actual infertile population (<0.4, 0.4–<2.8 and ≥2.8 ng/ml). Statistical analyses involved univariate comparisons between unsuccessful cycles and those that resulted in live birth. Since live birth may be influenced by multiple factors, a second set of analysis was performed using multivariate analysis. The multivariate logistic regression analysis with a stepwise backward selection procedure was used to develop a prediction model for the occurrence of the live birth using Wald $P < 0.05$ for entry and $P > 0.1$ for removal.

The probability of each live birth is given by the equation $P = 1/(1 + e^{(-p)})$ where $e^{(-p)}$ indicates that the base of the natural logarithm (2.718) is taken to the power of p, in which p is a linear formula derived from the regression coefficient of the significant variables. To reduce the overfit of the developed model, validation was performed by bootstrapping, hence adjusting the calculated model. Internal validation with bootstrapping (200 repetitions) was used to reduce the overfit of the model and to obtain relatively unbiased estimates. The bootstrap is a general data-based computational tool that can be used to assign measures of accuracy to statistical estimates (Steyerberg et al., 2001). The same multivariate logistic regression analysis was performed in the 200 data sets and a shrinkage factor was calculated by analysing the variability of the models. It was used to correct the final model and the prediction formula was extracted from the data. To evaluate the discrimination of the model the area under the receiver operating characteristic (ROC) curve was calculated. Pearson's chi-squared goodness of fit test was used to assess the overall performance of the model. Statistical analysis was performed by an accredited statistician of the University of Modena (RD) by using the software Stata 10 (StataCorp, TX, USA).

Results

A total of 389 patients were selected on the basis of inclusion criteria. Eight cycles were cancelled because of wrong drug administration, hence 381 patients constituted the population included in the statistical analysis. Of 381 started cycles, 15 were cancelled during ovarian stimulation because of excessive ovarian response and 13 because of absent or insufficient ovarian response. Of the 353 patients who had an oocyte retrieval, three had no oocytes retrieved and three had no fertilization; consequently, 347 patients had an embryo transfer procedure.

Baseline and cycle characteristics of patients are described in Tables 1 and 2, respectively. As expected, the main indications for treatment were unexplained and moderate male factors, with more than one cause identified in 17.5% of couples.

Of the cohort, 101 of 381 women (26.5%) achieved a live birth. Univariate and multivariate logistic regression was used to examine the association of live birth with baseline patient characteristics, in particular age, AMH, BMI and type, duration and aetiology of infertility. Univariate analysis revealed statistically significant decreasing odds of live birth for increasing age and decreasing AMH irrespective of whether they were treated as continuous or predestined categorical variables (Table 3). No association with live birth was observed for BMI or duration, type or cause of infertility. Analyses in multiple logistic regression models incorporating all predictors confirmed the independence of age and AMH in the prediction of live birth (Table 3).

Table 1 Baseline patient characteristics.

Characteristic	*Study population (n = 381)*
Age (years)	34.8 ± 4.48
BMI (kg/m^2)	24 ± 5.8
AMH (ng/ml)	1.3 (0.03–13.8)
Duration of infertility (months)	34.1 ± 20.2
Type of infertility	
Primary	294 (77.2)
Secondary	87 (22.8)
Cause of infertility	
Anovulation	82 (21.5)
Tubal factor	57 (15.0)
Unexplained	140 (36.8)
Male infertility	123 (32.4)
Endometriosis	45 (11.8)

Values are mean ± SD, median (range) or *n* (%).
AMH = anti-Müllerian hormone; BMI = body mass index.

The analysis indicated that women in the age categories 31–37 and >37 had a chance of live birth decreased by 39% and 64%, respectively, when compared with women younger than 31 years. Similarly women with AMH concentrations of 0.4–<2.8 ng/ml and <0.4 ng/ml had a chance of live birth decreased by 44% and 86%, respectively, when compared with women with AMH concentrations ≥2.8 ng/ml. To examine the validity of the regression model, the analysis was repeated on randomly selected subsamples and the reported results were confirmed by significance. Given that only AMH and age were demonstrated in univariate and multivariate analysis to predict live birth, a logistic regression model solely based on these two criteria was generated.

The probability for live birth depending on AMH and age can be calculated by the formula:

$$P(\text{live birth}) = \frac{\exp(-2.88 + 1.38 * \ln AMH_{1-2} + 1.96 * \ln AMH_3 + 1.01 * age_{<31} + 0.52 * age_{31-36})}{1 + \exp(-2.88 + 1.38 * \ln AMH_{1-2} + 1.96 * \ln AMH_3 + 1.01 * age_{<31} + 0.52 * age_{31-36})}$$

In order to evaluate whether the correlation existing between AMH and live birth was explained by the correlation existing between AMH and the number of retrieved oocytes, this was analysed by univariate and multivariate logistic regression. Although univariate analysis revealed statistically significant increasing odds of live birth for increasing number of oocytes (odds ration (OR) 1.06, 95% confidence intervals (CI) 1.01–1.1, $P < 0.05$), this was not statistically significant in the multivariate analyses and was therefore excluded from the model.

The discrimination ability of the model was assessed by determining the area under the ROC curve (AUC) and was 0.66 (95% CI 0.61–0.72) (**Figure 1**), which was significantly higher than the ROC curves of both AMH and age (ROC_{AUC} AMH 0.57, 95% CI 0.52–0.61, $P < 0.05$ and ROC_{AUC} age 0.55, 95% CI 0.52–0.59). At the best cut-off, the model permitted the identification of live birth with a sensitivity of 79.2%, specificity of 44.2% and the patients correctly classified were 53.5% (likelihood ratio, LR+ 1.42, LR− 0.46).

Assessment of the fit of the model was undertaken by Pearson's chi-squared, which confirmed that the model fitted well ($P = 0.55$). Table 4 demonstrates the predicted chance of a live birth versus observed live birth, with the difference in predicted and observed being <0.5%, which indicates a good calibration of the predictive model.

In order to facilitate the practical use of this model, a single three × three table was developed. By cross-tabulating any given patient's age with their AMH concentration, the probability of the live birth following IVF may be easily calculated (Table 5). According to the table, a woman aged 31–37 years and with serum AMH of 0.4–2.8 ng/ml has a 27% probability of achieving a live birth with a CI varying from 0.21 to 0.35.

Discussion

A number of factors have been reported as influencing the success of IVF either positively or negatively. A model which incorporates accurate estimates of the strength and independence of each factor in increasing or decreasing live birth rate would inevitably improve the advice underlying patient counselling on the basis of individualisation of likelihood of success. Furthermore, identification of patient characteristics which are directly linked with IVF outcome should enable individualisation of treatment strategies ensuring optimal outcomes even in first treatment cycles. To date, several models for the prediction of pregnancy after IVF have been proposed; however, few have been externally validated (Hunault et al., 2002; Stolwijk et al., 1996; Templeton et al., 1996) and only two have reported on live birth (Minaretzis et al. 1998; Templeton et al., 1996). With respect to models examining characteristics of patients which were available prior to treatment, female age has consistently been shown to be associated with IVF success. All other characteristics such as BMI, type of infertility (primary or secondary) and the diagnosis underlying the infertility have been found to be predictive of success in some but not all predictive models. Maternal age has also been strongly associated with outcome, even in models where the outcome of that first cycle, including oocyte yield, has been available for analysis. In accordance with these previous models, it is demonstrated here that among basal anamnestic characteristics only female age is strongly associated with live birth and warrants inclusion in the predictive model of live birth following IVF. No associations with cause, duration or type of infertility were observed. Although these observations differ from much historical data, which may reflect the limited size of the cohort, it may also indicate an overall improvement in treatment success rates with use of ICSI for male factor and optimisation of concurrent medical conditions prior to the treatment cycle. Consistent with this there is a marked improvement in overall live birth success rates in the current study despite the limits on oocyte insemination compared with the major historical series (26.5% versus 13.9%; Templeton et al., 1996).

The current study clearly demonstrates that AMH, a biomarker of ovarian reserve, is significantly associated with live birth and that this association is independent of age. Furthermore, it is demonstrated that a composite model of AMH and age can predict the probability of live birth following the first IVF/ICSI treatment cycle. Notably AMH was measured on stored samples taken prior to the IVF cycle and therefore knowledge of AMH values could not have altered clinical management and thereby model performance. Several existing studies have investigated the role of markers of ovarian reserve in the prediction of IVF success. Many of these studies frequently adopted FSH as a marker of ovarian reserve, globally reporting a positive role for its inclusion in the model (Bancsi et al., 2000; Ferlitsch et al., 2004; Ottosen et al., 2007; Younis et al., 2009). As far as is known, only a few studies have positively investigated other markers of ovarian reserve such as antral follicle count (Carrera-Rotllan et al., 2007; Lee et al., 2009; Maseelall et al., 2009; Younis et al., 2009) or AMH (Lee et al., 2009; Nelson et al., 2007). Despite the heterogeneity in choice of marker, all studies have concluded that inclusion of an ovarian reserve marker may improve the prediction of live birth. Notably, AMH is a better predictor of response to ovarian stimulation than FSH and equivalent to that of antral follicle count, and therefore may be an optimal secondary basal characteristic for the prediction of live birth.

Table 2 Outcome characteristics of the patient cohort.

Characteristic	*Study population (n = 381)*
Age at stimulation (years)	34.8 ± 4.48
Duration of stimulation (days)	12.8 ± 2.8
Total FSH dose (IU)	2624 ± 750
Oocytes per patient[a]	8.5 ± 5.1
Inseminated oocytes per patient	2.8 ± 0.59
Embryo transfers performed	347 (91.1)
Embryos transferred	
0[b]	34
1	8
2	83
3	256
Clinical pregnancies per started cycle	127 (33.3)
Live births per started cycle	101 (26.5)

Values are presented as mean ± SD or *n* (%).
[a]Patients reaching oocyte retrieval = 353.
[b]No transfer includes women who either had the cycle cancelled due to failure of response to gonadotrophin (*n* = 13), excessive response (*n* = 15), no oocytes at the retrieval (*n* = 3) or no fertilization (*n* = 3).

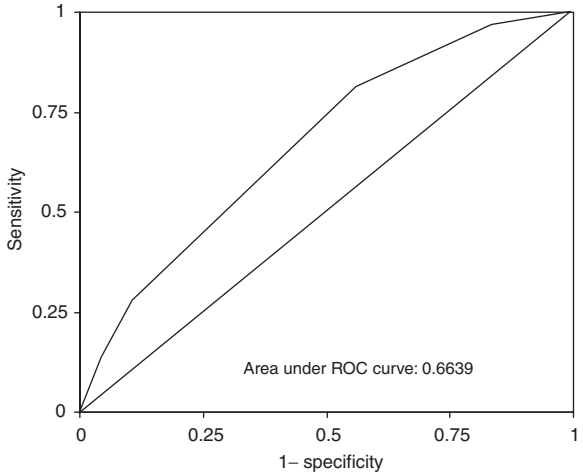

Figure 1 Receiver operating characteristic curve for the live birth prediction model. The discrimination ability of the model was assessed by determining the area under the curve and was 0.66 (95% CI 0.61 to 0.72). At the best cut-off, the model permitted the identification of live birth with a sensitivity of 79.2%, specificity of 44.2% and the patients correctly classified were 53.5%.

Regarding AMH, the prediction of qualitative aspects of assisted reproduction programmes by its measurement is largely controversial. This is also evident from studies reporting on the pregnancy rate following IVF. A number of authors have tried to identify an absolute concentration value for AMH that is able to distinguish between pregnancy and non-pregnancy (Elgindy et al., 2008; Eldar-Geva et al.,

Table 3 Association between live birth and predictive variables of treatment outcome measured at baseline.

Variables	*Live births per patient*	*Univariate OR (95% CI)*	*P-value*	*Multivariate OR (95% CI)*
Age (years)				
<31	30/70 (42.9)	Reference	0.0002	Reference
31–37	48/172 (27.9)	0.52 (0.29–0.92)		0.61 (0.34–1-12)
>37	23/139 (16.5)	0.26 (0.14–0.51)		0.36 (0.18–0.72)
BMI (kg/m^2)				
<26	60/224 (26.8)	Reference	NS	–
26–30	32/120 (26.7)	1 (0.99–1.00)		
>30	9/37 (24.3)	0.88 (0.6–1.829		
AMH (ng/ml)				
≥2.8	29/67 (43.3)	Reference	0.00004	Reference
0.4–<2.8	69/270 (25.6)	0.45 (0.26–0.78)		0.56 (0.31–0.99)
<0.4	3/44 (6.8)	0.1 (0.03–0.349)		0.14 (0.04–0.51)
Duration of infertility		0.99 (0.98–1.00)	NS	–
Type of infertility				
Primary	75/294 (25.5)	Reference	NS	–
Secondary	26/87 (29.9)	1.24 (0.73–2.11)		
Cause of infertility				
Anovulation	22/82 (26.8)	1.02 (0.59–1.77)	NS	–
Tubal factor	14/57 (24.6)	0.89 (0.46–1.70)	NS	–
Unexplained	39/140 (27.9)	1.11 (0.70–1.78)	NS	–
Male infertility	32/123 (26.0)	0.98 (0.60–1.59)	NS	–
Endometriosis	9/45 (20.0)	0.66 (0.31–1.43)	NS	–

Values are *n*/total (%).
Univariate analysis revealed statistically significant decreasing odds of live birth for increasing age and decreasing AMH irrespective of whether they were treated as continuous or predestined categorical variables.
AMH = anti-Müllerian hormone; BMI = body mass index; NS = not statistically significant.

Table 4 Expected chance of a live birth versus observed live birth, with the difference in predicted and observed <0.5% indicating a good calibration of the predictive model (Pearson chi-squared goodness-of-fit test).

Covariate patterns	*Probability*[a]	*Live birth*		*No live birth*		*Total*
		Observed	*Expected*	*Observed*	*Expected*	
1	0.0531	1	1.4	26	25.6	27
2	0.0865	1	1.3	14	13.7	15
3	0.1336	1	0.3	1	1.7	2
4	0.1824	18	18.4	83	82.6	101
5	0.2734	36	35.0	92	93.0	128
6	0.2861	4	3.1	7	7.9	11
7	0.3800	15	15.6	26	25.4	41
8	0.4035	11	11.7	18	17.3	29
9	0.5242	14	14.2	13	12.8	27

Number of observations = 381; Pearson chi-squared (4) = 3.02; probability > chi-squared = 0.5552.
[a]Probability of live birth for the specific covariate pattern.

Table 5 Probability (95% CI) of live birth after IVF according to age and AMH.

Age (years)	*AMH (ng/ml)*		
	<0.4	*0.4–<2.8*	*≥2.8*
<31	0.13 (0.04–0.36)	0.38 (0.26–0.51)	0.52 (0.38–0.67)
31–37	0.09 (0.02–0.24)	0.27 (0.21–0.35)	0.40 (0.28–0.54)
>37	0.05 (0.01–0.16)	0.18 (0.12–0.26)	0.29 (0.17–0.44)

Probability of live birth was obtained by using the parameters estimated from the logistic model:

$$P(\text{live birth}) = \frac{\exp(-2.88 + 1.38 * \ln AMH_{1-2} + 1.96 * \ln AMH_{3} + 1.01 * age_{<31} + 0.52 * age_{31-36})}{1 + \exp(-2.88 + 1.38 * \ln AMH_{1-2} + 1.96 * \ln AMH_{3} + 1.01 * age_{<31} + 0.52 * age_{31-36})}$$

2005; Hazout et al., 2004; Kwee et al., 2008; Nelson et al., 2009a). However, the majority of them indicated that AMH measurement is not useful for predicting this end-point (Ebner et al., 2006; Fanchin et al., 2003; Fiçicioglu et al., 2006; Kwee et al., 2008; Nelson et al., 2009b; Peñarrubia et al., 2005; Smeenk et al., 2007; Van Rooij et al., 2002). Consistent with a non-discriminative point, a large prospective cohort study of 340 patients, relating serum AMH concentrations to the live birth rate following IVF, demonstrated a positive association of live birth and basal AMH, with improved predictability compared with basal FSH (Nelson et al., 2007). Furthermore, a very recent prospective study of 336 patients undergoing their first IVF cycle has clearly demonstrated that among the ovarian reserve tests, AMH and female age had a greater area under the ROC curve than FSH in predicting live birth (Lee et al., 2009). In that cohort, AMH and age were the sole predictive factors of live birth for women ≥ 35 years, with only the number of good-quality embryos predicting live birth in women < 35 (Lee et al., 2009). However, it is important to note that prior to stimulation only AMH and age would be available to counsel patients, and that AMH and oocyte yield, and thereby embryo number, are intrinsically linked (La Marca et al., 2009). The present study therefore has substantial benefits as it demonstrates a strong predictive performance of AMH for live birth at all female ages, permitting the construction of a model based on only two parameters, namely age and serum AMH concentrations. Furthermore, analysis of the goodness of fit-test (Table 4) demonstrated that the model correctly fits the data.

A further substantive difference between the current study and that by Lee et al. (2009) is the number of transferred embryos, with Lee reporting a mean of 3.9 embryos transferred in women aged <35 and 3.1 in women aged ≥35, reflecting differences in policy between Taiwan and Italy. Of course, young women produce a high number of oocytes and embryos, hence permitting selection of embryos for transfer. The transfer of a high number of good-quality embryos may be sufficient to overcome a possible influence of any factor on the success of IVF. This seems to be demonstrated by the fact that in the study by Lee, the live birth in women <35 is predicted only by the number of good-quality transferred embryos whereas AMH and age predicted live birth in older women (for whom a reduced number of embryos is usually available). More importantly, the current study has been performed according to the Italian law

regulating assisted reproduction, which limits the number of inseminated oocytes to three and thus reduces the number of embryos that may be generated for each patient. This of course is a limitation for young women who generally produce a high number of oocytes and for whom selection of the optimal embryos for transfer cannot be performed. This particular policy, however, has permitted for the first time to gain information regarding success of IVF independent of the selection of embryos. Consequently in this study, AMH has been shown to predict the chance of success in both younger and older women. Interestingly, analysis of whether AMH predicted live birth independently of a correlation with oocyte yield was performed by inclusion of oocyte yield in the model. When the number of retrieved oocytes was excluded from the model, AMH and age were the only predictors of live birth. This suggests that AMH may somehow be linked not only to the quantity but also to quality of female gametes.

The area under the ROC curve of the model was found to be significantly higher than either age or AMH alone. However it should be highlighted that the model does not seem to have an optimal discriminative performance (ROC_{AUC} 0.66). In practical terms, this may mean that differences are not large enough for practitioners to counsel their patients in terms of chance for live birth. It should be emphasized that the predictive power is relatively limited. This can be explained by the fact that the spectrum of disease is narrow in couples undergoing IVF, i.e. the test sample includes a strong overlap between couples who conceive and couples who failed to conceive. Rather than ROC curves which are primarily designed for diagnostic models (Cook (2008)), the predictive accuracy of a prognostic model can be expressed by calibration and discrimination (Harrel et al., 1996). Discrimination does not reflect the accuracy of a model and its clinical significance is poor. Calibration on the other hand is an important parameter reflecting the accuracy of prediction. Calibration is evaluated by accessing the level of correspondence between the calculated pregnancy probabilities and the observed proportion of pregnancies. Well-calibrated models are able to classify individuals into clinically useful prognostic strata on the basis of the calculated probabilities of a pregnancy with or without treatment. This is illustrated by the external validation of the Templeton model (Smeenk et al., 2000) for the prediction of pregnancy after IVF. The model differentiates between couples with low and high probability of success despite its limited discrimination between couples with or without success (Smeenk et al., 2000). In contrast to relatively poor discrimination, the calibration of the model was found to be good (Table 4). The importance of discrimination and calibration depends on the clinical application of the model. This model is intended to counsel couples, thus the accuracy of the numeric probability (calibration) is important. Patients are not concerned about how their chance is relative to other couples (discrimination); instead, they want to know the absolute likelihood that they will get pregnant within the IVF cycle. Consequently the clinical aim of the model is to differentiate between couples with either a poor or good prognosis.

The use of AMH may be proposed as a diagnostic test to inform the patients about their chance in assisted reproduction treatment, allowing adjustment to the already useful information derived from patient age. According to the model presented here, the patients with low basal AMH concentrations have a low chance of success, especially if they are older than 37 (predicted probability of live birth 5%).

If further substantiated, this information has extensive applications for patients making informed decisions, focusing adaptive research and also treatment funding bodies. Whether an expected live birth rate of 5% would be considered as acceptable in a public state-funded centre is not clear, but this information will allow the debate to take place on a firm footing. Conversely, older women with high AMH concentrations may be considered as having an enhanced qualitative and quantitative ovarian reserve. Furthermore, given that the predicted live birth rate is approximately two-fold higher than that for young women with a low AMH, provision of funding to this group would potentially be worthwhile and contrary to current dogma where age primarily dictates access to assisted conception services.

Although it is practically impossible to predict the individual chance of a live birth in a couple accurately, prognostic models can help to address these matters in a more objective way. They can also act as a convincing tool in individual counselling for both patients as well as physicians. However, it remains to be seen how many patients refrain from treatment if their prognostic chance is poor. Before the model can be used in the clinical setting, external validation should be performed. Although external validation of some prediction models have resulted in a lower predictive performance (Stolwijk et al., 1998), others have demonstrated a good predictive potential in other populations (Hunault et al., 2007).

In conclusion, the present study demonstrates that in the prediction of a live birth following IVF, a distinction, however moderate, can be made between couples with a good and a poor prognosis. The success of IVF was found to mainly depend on maternal age and serum AMH concentrations, one of the most relevant and valuable markers of ovarian reserve.

Acknowledgements

The authors thank Professor Richard Fleming for his constructive comments on a draft of the manuscript.

References

Bancsi, L.F., Huijs, A.M., den Ouden, C.T., et al., 2000. Basal follicle-stimulating hormone levels are of limited value in predicting ongoing pregnancy rates after in vitro fertilization. Fertil. Steril. 73, 552—557.

Bouckaert, A., Psalti, I., Loumaye, E., De Cooman, S., Thomas, K., 1994. The probability of a successful treatment of infertility by in-vitro fertilization. Hum. Reprod. 9, 448—455.

Carrera-Rotllan, J., Estrada-García, L., Sarquella-Ventura, J., 2007. Prediction of pregnancy in IVF cycles on the fourth day of ovarian stimulation. J. Assist. Reprod. Genet. 24, 387—394.

Commenges-Ducos, M., Tricaud, S., Papaxanthos-Roche, A., Dallay, D., Horovitz, J., Commenges, D., 1998. Modelling of the probability of success of the stages of in-vitro fertilization and embryo transfer: stimulation, fertilization and implantation. Hum. Reprod. 13, 78—83.

Cook, N.R., 2008. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. Clin. Chem. 54, 17–23.

Cupisti, S., Dittrich, R., Mueller, A., et al., 2007. Correlations between anti-Müllerian hormone, inhibin B, and activin A in follicular fluid in IVF/ICSI patients for assessing the maturation and developmental potential of oocytes. Eur. J. Med. Res. 12, 604–608.

Ebner, T., Sommergruber, M., Moser, M., Shebl, O., Schreier-Lechner, E., Tews, G., 2006. Basal level of anti-Müllerian hormone is associated with oocyte quality in stimulated cycles. Hum. Reprod. 21, 2022–2026.

Eldar-Geva, T., Ben-Chetrit, A., Spitz, I.M., et al., 2005. Dynamic assays of inhibin B, anti-Mullerian hormone and estradiol following FSH stimulation and ovarian ultrasonography as predictors of IVF outcome. Hum. Reprod. 20, 3178–3183.

Elgindy, E.A., El-Haieg, D.O., El-Sebaey, A., 2008. Anti-Müllerian hormone: correlation of early follicular, ovulatory and midluteal levels with ovarian response and cycle outcome in intracytoplasmic sperm injection patients. Fertil. Steril. 89, 1670–2167.

Fanchin, R., Schonauer, L.M., Righini, C., Guibourdenche, J., Frydman, R., Taieb, J., 2003. Serum anti-Mullerian hormone is more strongly related to ovarian follicular status than serum inhibin B, estradiol, FSH and LH on day 3. Hum. Reprod. 18, 323–327.

Ferlitsch, K., Sator, M.O., Gruber, D.M., Rücklinger, E., Gruber, C.J., Huber, J.C., 2004. Body mass index, follicle-stimulating hormone and their predictive value in in vitro fertilization. J. Assist. Reprod. Genet. 21, 431–436.

Fiçicioglu, C., Kutlu, T., Baglam, E., Bakacak, Z., 2006. Early follicular anti-Müllerian hormone as an indicator of ovarian reserve. Fertil. Steril. 85, 592–596.

Haan, G., Bernardus, R.E., Hollanders, J.M., Leerentveld, R.A., Prak, F.M., Naaktgeboren, N., 1991. Results of IVF from a prospective multicentre study. Hum. Reprod. 6, 805–810.

Harrell Jr, F.E., Lee, K.L., Mark, D.B., 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat. Med. 15, 361–387.

Hazout, A., Bouchard, P., Seifer, D.B., Aussage, P., Junca, A.M., Cohen-Bacrie, P., 2004. Serum antimullerian hormone/mullerian-inhibiting substance appears to be a more discriminatory marker of assisted reproductive technology outcome than follicle-stimulating hormone, inhibin B, or estradiol. Fertil. Steril. 82, 1323–1329.

Hughes, E.G., King, C., Wood, E.C., 1998. A prospective study of prognostic factors in in vitro fertilization and embryo transfer. Fertil. Steril. 51, 838–844.

Hull, M.G., Fleming, C.F., Hughes, A.O., McDermott, A., 1996. The age-related decline in female fecundity: a quantitative controlled study of implanting capacity and survival of individual embryos after in vitro fertilization. Fertil. Steril. 65, 783–790.

Hunault, C.C., Eijkemans, M.J., Pieters, M.H., et al., 2002. A prediction model for selecting patients undergoing in vitro fertilization for elective single embryo transfer. Fertil. Steril. 77, 725–732.

Hunault, C.C., te Velde, E.R., Weima, S.M., et al., 2007. A case study of the applicability of a prediction model for the selection of patients undergoing in vitro fertilization for single embryo transfer in another center. Fertil. Steril. 87, 1314–1321.

Kwee, J., Schats, R., McDonnell, J., Themmen, A., de Jong, F., Lambalk, C., 2008. Evaluation of anti-Müllerian hormone as a test for the prediction of ovarian reserve. Fertil. Steril. 90, 737–743.

La Marca, A., Sighinolfi, G., Radi, D., Argento, C. et al., 2009. Anti-Mullerian hormone (AMH) as a predictive marker in assisted reproductive technology (ART). Hum. Reprod. Update (Epub ahead of print).

Lee, T.H., Liu, C.H., Huang, C.C., Hsieh, K.C., Lin, P.M., Lee, M.S., 2009. Impact of female age and male infertility on ovarian reserve markers to predict outcome of assisted reproduction technology cycles. Reprod. Biol. Endocrinol. 7, 100.

Lekamge, D.N., 2007. Anti-Mullerian hormone as a predictor of IVF outcome. RBMonline 14, 602–610.

Lie Fong, S., Baart, E.B., Martini, E., et al., 2008. Anti-Müllerian hormone: a marker for oocyte quantity, oocyte quality and embryo quality? Reprod. Biomed. Online 16, 664–670.

Lintsen, A.M., Pasker-de Jong, P.C., de Boer, E.J., et al., 2005. Effects of subfertility cause, smoking and body weight on the success rate of IVF. Hum. Reprod. 20, 1867–1875.

Lintsen, A.M., Eijkemans, M.J., Hunault, C.C., et al., 2007. Predicting ongoing pregnancy chances after IVF and ICSI: a national prospective study. Hum. Reprod. 22, 2455–2462.

Maseelall, P.B., Hernandez-Rey, A.E., Oh, C., Maagdenberg, T., McCulloh, D.H., McGovern, P.G., 2009. Antral follicle count is a significant predictor of livebirth in in vitro fertilization cycles. Fertil. Steril. 91, 1595–1597.

Minaretzis, D., Harris, D., Alper, M.M., Mortola, J.F., Berger, M.J., Power, D., 1998. Multivariate analysis of factors predictive of successful live births in in vitro fertilization (IVF) suggests strategies to improve IVF outcome. J. Assist. Reprod. Genet. 15, 365–371.

Nelson, S.M., Yates, R.W., Fleming, R., 2007. Serum anti-Müllerian hormone and FSH: prediction of live birth and extremes of response in stimulated cycles—implications for individualization of therapy. Hum. Reprod. 22, 2414–2421.

Nelson, S.M., Yates, R.W., Lyall, H., et al., 2009a. Anti-Müllerian hormone-based approach to controlled ovarian stimulation for assisted conception. Hum. Reprod. 24, 867–875.

Nelson, S.M., Fleming, R., 2009b. Low AMH and GnRH-antagonist strategies. Fertil. Steril. 92, e40.

Ottosen, L.D., Kesmodel, U., Hindkjaer, J., Ingerslev, H.J., 2007. Pregnancy prediction models and eSET criteria for IVF patients–do we need more information? J. Assist. Reprod. Genet. 24, 29–36.

Peñarrubia, J., Fábregues, F., Manau, D., et al., 2005. Basal and stimulation day 5 anti-Mullerian hormone serum concentrations as predictors of ovarian response and pregnancy in assisted reproductive technology cycles stimulated with gonadotropin-releasing hormone agonist—gonadotropin treatment. Hum. Reprod. 20, 915–922.

Silberstein, T., MacLaughlin, D.T., Shai, I., et al., 2006. Mullerian inhibiting substance levels at the time of HCG administration in IVF cycles predict both ovarian reserve and embryo morphology. Hum. Reprod. 21, 159–163.

Smeenk, J.M., Stolwijk, A.M., Kremer, J.A., Braat, D.D., 2000. External validation of the templeton model for predicting success after IVF. Hum. Reprod. 15, 1065–1068.

Smeenk, J.M., Sweep, F.C., Zielhuis, G.A., Kremer, J.A., Thomas, C.M., Braat, D.D., 2007. Antimüllerian hormone predicts ovarian responsiveness, but not embryo quality or pregnancy, after in vitro fertilization or intracyoplasmic sperm injection. Fertil. Steril. 87, 223–226.

Steyerberg, E.W., Harrell Jr., F.E., Borsboom, G.J., Eijkemans, M.J., Vergouwe, Y., Habbema, J.D., 2001. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J. Clin. Epidemiol. 54, 774–781.

Stolwijk, A.M., Zielhuis, G.A., Hamilton, C.J., et al., 1996. Prognostic models for the probability of achieving an ongoing pregnancy after in-vitro fertilization and the importance of testing their predictive value. Hum. Reprod. 11, 2298–2303.

Stolwijk, A.M., Straatman, H., Zielhuis, G.A., et al., 1998. External validation of prognostic models for ongoing pregnancy after in-vitro fertilization. Hum. Reprod. 13, 3542–3549.

Stolwijk, A.M., Wetzels, A.M., Braat, D.D., 2000. Cumulative probability of achieving an ongoing pregnancy after in-vitro fertilization and intracytoplasmic sperm injection according to a woman's age, subfertility diagnosis and primary or secondary subfertility. Hum. Reprod. 15, 203—209.

te Velde, E.R., Pearson, P.L., 2002. The variability of female reproductive ageing. Hum. Reprod. Update 8, 141—154.

Templeton, A., Morris, J.K., Parslow, W., 1996. Factors that affect outcome of in-vitro fertilisation treatment. Lancet, 3481402—3481406.

Van Rooij, I.A., Broekmans, F.J., te Velde, E.R., et al., 2002. Serum anti-Mullerian hormone levels: a novel measure of ovarian reserve. Hum. Reprod. 17, 3065—3071.

van Weert, J.M., Repping, S., van der Steeg, J.W., Steures, P., van der Veen, F., Mol, B.W., 2008. A prediction model for ongoing pregnancy after in vitro fertilization in couples with male subfertility. J. Reprod. Med. 53, 250—256.

Verberg, M.F., Eijkemans, M.J., Macklon, N.S., Heijnen, E.M., Fauser, B.C., Broekmans, F.J., 2008. Predictors of ongoing pregnancy after single-embryo transfer following mild ovarian stimulation for IVF. Fertil. Steril. 89, 1159—1165.

Verberg, M.F., Eijkemans, M.J., Macklon, N.S., Heijnen, E.M., Fauser, B.C., Broekmans, F.J., 2007. Predictors of low response to mild ovarian stimulation initiated on cycle day 5 for IVF. Hum. Reprod. 22, 1919—1924.

World Health Organization, 1999. Laboratory Manual for the Examination of Human Semen and Sperm-Cervical Mucus Interaction. WHO, fourth ed.

Younis, JS., Jadaon, J., Izhaki, I. et al., 2009. A simple multivariate score could predict ovarian reserve, as well as pregnancy rate, in infertile women. Fertil. Steril. (Epub ahead of print).

Declaration: The authors report no financial or commercial conflicts of interest.

Received 26 May 2010; refereed 2 November 2010; accepted 2 November 2010.