



The (decision) tree of fertility: an innovative decision-making algorithm in assisted reproduction technique

Maria Teresa Villani¹ · Daria Morini¹ · Giorgia Spaggiari² · Chiara Furini^{2,3} · Beatrice Melli^{1,4} · Alessia Nicoli¹ · Francesca Iannotti¹ · Giovanni Battista La Sala¹ · Manuela Simoni^{2,3} · Lorenzo Aguzzoli¹ · Daniele Santi^{2,3}

Received: 24 August 2021 / Accepted: 5 November 2021 / Published online: 27 January 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Purpose Several mathematical models have been developed to estimate individualized chances of assisted reproduction techniques (ART) success, although with limited clinical application. Our study aimed to develop a decisional algorithm able to predict pregnancy and live birth rates after controlled ovarian stimulation (COS) phase, helping the physician to decide whether to perform oocytes pick-up continuing the ongoing ART path.

Methods A single-center retrospective analysis of real-world data was carried out including all fresh ART cycles performed in 1998–2020. Baseline characteristics, ART parameters and biochemical/clinical pregnancies and live birth rates were collected. A seven-steps systematic approach for model development, combining linear regression analyses and decision trees (DT), was applied for biochemical, clinical pregnancy, and live birth rates.

Results Of fresh ART cycles, 12,275 were included. Linear regression analyses highlighted a relationship between number of ovarian follicles > 17 mm detected at ultrasound before pick-up (OF17), embryos number and fertilization rate, and biochemical and clinical pregnancy rates ($p < 0.001$), but not live birth rate. DT were created for biochemical pregnancy (statistical power–SP:80.8%), clinical pregnancy (SP:85.4%), and live birth (SP:87.2%). Thresholds for OF17 entered in all DT, while sperm motility entered the biochemical pregnancy's model, and female age entered the clinical pregnancy and live birth DT. In case of OF17 < 3, the chance of conceiving was < 6% for all DT.

Conclusion A systematic approach allows to identify OF17, female age, and sperm motility as pre-retrieval predictors of ART outcome, possibly reducing the socio-economic burden of ART failure, allowing the clinician to perform or not the oocytes pick-up.

Keywords Assisted reproduction · Controlled ovarian stimulation · Oocytes · Decision tree · Ovarian follicles

Maria Teresa Villani and Daria Morini have contributed equally to this work and share first authorship.

✉ Giorgia Spaggiari
spaggiari.giorgia@aou.mo.it

¹ Department of Obstetrics and Gynaecology, Fertility Centre, Azienda Unità Sanitaria Locale-IRCCS di Reggio Emilia, Arcispedale Santa Maria Nuova, Reggio Emilia, Italy

² Unit of Endocrinology, Department of Medical Specialties, Azienda Ospedaliero-Universitaria di Modena, Ospedale Civile di Baggiovara, Via Giardini 1355, 41126 Modena, Italy

³ Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy

⁴ Clinical and Experimental Medicine PhD Program, University of Modena and Reggio Emilia, Modena, Italy

Introduction

Infertility, defined as the inability to achieve clinical pregnancy after 1 year of regular unprotected sexual intercourse, is estimated to affect between 8 and 16% of reproductive-aged couples worldwide [1]. Thus, unsurprisingly, assisted reproduction techniques (ART) are increasingly applied in current clinical practice, due both to the relevant incidence of couple infertility and to the advanced age of couples starting to search a pregnancy [2]. Globally, it is estimated that more than eight million babies have been conceived through ART [3]. However, infertility treatment is a long-term and expensive therapy with high dropout rates [4] and nearly half of all couples who started ART are likely to remain childless even after multiple treatment cycles, with foreseeable sequelae in terms

of psychological, social, and economic health [5–7]. In this context, mathematical models are generated predict strong outcomes, such as pregnancy and live birth rates. On these parameters, several predictive models for ART have been developed over the last three decades in order to estimate individualized chances of treatment success [8]. These models are required to select either the ART approach applicable to the couple or the best treatment for the female partner [9]. However, these algorithms are scanty applied in current clinical practice. Nowadays, scientific societies suggest to tailor COS schemes on the female characteristics [10–12].

In clinical practice, the physician needs to estimate a priori the female response after the controlled stimulation phase (COS) without clear evidence-based recommendations [13], leaving an extreme variability in the proposed therapeutic regimens [8]. Indeed, the most cost-effective ART management in terms of pregnancy and live birth rates is still far to be achieved [14]. Alongside the absence of a gold standard of care in ART, the clinical application of predictive models is still scanty, considering their limited predictive ability and lack of confidence among clinicians about their effectiveness [15].

Despite the wide literature describing how to perform a predictive research, the majority of models published so far suffer from methodological shortcomings [16, 17]. One of the most accredited predictive models is the “McLernon post-treatment model,” which predicts the cumulative probability of a live birth after the first fresh embryo transfer during one or more in vitro fertilization (IVF) or intracytoplasmic sperm injection (ICSI) cycles [18]. Within this post-treatment model, the following parameters were identified as best predictors for live birth rate: (i) woman’s age, (ii) number of oocytes retrieved, and (iii) cryopreservation of embryos [18]. However, the currently available predictive models are not sufficiently reliable to guarantee uniform counseling for infertile couples [8]. Moreover, these parameters are generally available only when the ART path has been started and has passed a point of no return, which cannot be changed. Indeed, knowing that the chances of ART success are low or equal to zero when the number of embryos is low, does not allow the clinician to change the approach or suspend treatment before failure, with the known psychological and economic consequences on the couple.

Thus, the need to obtain reliable parameters able to improve the concordance of treatment decisions in reproductive medicine remains urgent [8]. In particular, to promote the clinical impact of predictive models, it could be useful to identify predictors for those ART phases which could be revised or interrupted, i.e., COS, pick-up and embryo transfer phases. In particular, a predictive model able to estimate the chances of success in the time point after ovarian stimulation and before the pick-up could guide the decision

to prosecute or not the ongoing ART path or to suspended it and to re-schedule a new COS.

With this in mind, the aim of this study is the development of a decisional algorithm able to predict strong ART outcomes, i.e., pregnancy and live birth rates, in order to help the clinician to decide when and whether to perform oocytes pick-up, continuing the ongoing ART path.

Materials and methods

A single center, retrospective analysis of real-world data was carried out, considering all couples attending the Fertility Centre of the Department of Obstetrics and Gynaecology of Reggio Emilia (Italy).

All consecutive ART cycles performed from 1998 to December 2020 were retrospectively extracted and couples fulfilling following inclusion criteria were included in the final dataset. Couple with (i) both partners older than 18 years, (ii) attending ART cycles performed using fresh sperms and oocytes, (iii) in which COS proceeded until ovulation, and (iv) all ART variables reported below are available. Thus, neither cycles stopped for any reasons, nor cycles performed with frozen sperms and/or oocytes have been included. Donor egg and donor sperm were excluded. Both ICSI and IVF cycles were considered.

Assisted reproductive technology (ART) procedures

The downregulation of the hypothalamic–pituitary–gonadal axis was obtained through gonadotropin-releasing hormone agonists administration (GnRHa) (Enantone®, Takeda Pharmaceutical, or Decapeptyl®, Ipsen). Then, ovarian stimulation was performed applying individualized protocols: (i) recombinant follicle-stimulating hormone (FSH) alone (Gonal-F®, Merck Serono), (ii) recombinant FSH plus luteinizing hormone (LH) (Pergoveris®, Merck Serono), (iii) highly purified human menopausal gonadotropin (hMG) (Meropur®, Ferring), or (iv) biosimilar FSH (Ovaleap®, Theramex). The ovarian stimulation was monitored by serum estradiol assays and serial ultrasound (US) evaluations. When more than three follicles with diameter higher than 17 mm were observed at US, human chorionic gonadotropin (hCG) (Gonasi®, IBSA Institut Biochimique) was injected to complete oocyte maturation and to promote ovulation. The oocyte retrieval was performed 34–36 h after hCG administration by US-guided transvaginal aspiration. All patients received supplemental progesterone for 12 days until β -hCG assay.

For conventional IVF, oocytes were individually cultured in microdrops of fresh medium under mineral oil with 100,000 activated sperms. For ICSI, after the removal of the cumulus and corona cells, nuclear maturation assessment

of oocytes was performed using an inverted microscope to ensure the sperm injection in metaphase-II oocytes only.

Oocytes fertilization was assessed at 18–20 h (day 1) after insemination/injection and confirmed by the presence of two pronuclei and the alignment of nucleolar precursor bodies. In all cases, the embryonic development was assessed on days 2 and 3 (i.e., after 41–43 and 65–67 h from insemination/injection, respectively). The best-quality embryos were transferred on day 2 or 3 after IVF/ICSI procedures until July 2020, when blastocyst transfer was started. In Italy, embryo production and transfer are regulated by specific national laws which have changed over the years. In particular, until 2004, a maximum of five embryos were transferred for each cycle, since embryos freezing was not allowed. Afterwards, the allowed maximum number of transferred embryos has progressively decreased until July 2020, when it was set to one embryo for women younger than 38 years and two for women older. Moreover, across years of observations, the day in which the embryo could be transferred changed according to national rules. Since the embryo transfer follows COS, the day in which it was performed was not considered as a predictive variable of the statistical analysis, limiting its potential bias.

For the evaluation of pregnancies, international ESHRE definitions were considered [19]. In particular, biochemical pregnancies were assessed 12 days after embryo transfer by a positive quantitative serum β -hCG assay higher than 10 IU/L. In case of positive biochemical pregnancy test, micronized progesterone support (Prometrium®, Rottapharm Madaus or Crinone®, Merck Serono) was continued until 35 days after embryo transfer.

Outcomes

Baseline couple characteristics were collected, considering age, body mass index (BMI), and smoking habit of both partners. Moreover, reason of couple infertility and fertility history of the couple (previous pregnancies, miscarriages, pre-term, and term births) were collected as categorical data.

The ART procedure was evaluated collecting several variables, considering male parameters (e.g., semen volume, sperm concentration, sperm motility, and morphology percentages), COS approach (e.g., gonadotropin drug used, days of stimulation, starting gonadotropin dosage, total gonadotropins units used), and variables of COS response (e.g., ovarian follicles > 17 mm detected at US before pick-up—OF17, total and mature oocytes retrieved, injected/inseminated oocytes, fertilized oocytes, fertilization rate and number of total, transferred, and frozen embryos). The fertilization rate was calculated a posteriori as the ratio between the number of fertilized oocytes and the number of either injected (ICSI method) or inseminated (IVF cycles) oocytes.

Finally, the strong ART outcomes were considered, i.e., biochemical and clinical pregnancy and live birth rates. The biochemical pregnancy rate represented the detection of increased levels of hCG in serum, while the clinical pregnancy was diagnosed in case of US visualization of at least one embryos with heartbeat [19].

Statistical analysis

First, the entire dataset was evaluated performing descriptive statistics, in order to obtain a snapshot of the characteristics of the cohort included, evaluating both ART variables and outcomes.

Second, continuous parameters' distribution was evaluated by Kolmogorov–Smirnov test. Then, continuous data were compared between couples who obtained a pregnancy (biochemical and clinical separately) and a live birth, using either ANOVA univariate or Mann–Whitney *U*-test, according to data distribution.

Linear regression logistic analyses were performed, repeating the analysis for each strong ART outcome. In details, these analyses were performed using strong ART outcomes as dependent variables, setting both cohort baseline characteristics and ART variables reported above as covariates and cofactors. Among cofactors, the ART approach used (i.e., ICSI or IVF) and the couple infertility etiology were included. Only those ART variables which predicted pregnancy/live birth rates were extracted and evaluated in correlation analyses with other variables, using Spearman's Rho analysis. Moreover, these variables were used as dependent variables in multiple linear regression analyses, setting other ART variables, and baseline cohort characteristics as independent variables. These analyses are needed to decide whether a predictive model could be developed considering only variables obtained before pick-up.

The predictive models' development was systematically performed according to the seven steps proposed by Steyerberg et al. [20], considering separately biochemical pregnancy, clinical pregnancy, and live birth rates as final outcome. Table 1 shows the systematic approach to the main question.

Finally, three decision trees were created. Each decision tree was built to predict pregnancy (biochemical and clinic separately) and live birth rates, considering the study question. Thus, the dependent variables were the strong ART outcomes, whereas independent variables/factors were the ART variables that precede pick-up. The exhaustive chi-square automatic interaction detector (CHAID) decision tree was applied. This statistical tool derived from the first algorithm developed in the 1980s, accepting both categorical and continuous variables [21]. In particular, like CHAID, merely nominal or ordinal categorical predictors are allowable, thus continuous predictors are first

Table 1 Development of predictive models applying the seven-steps systematic approach

| | Question | Answer |
|--------------|--|--|
| First step | Problem definition and data inspection | Main goal of the predictive model: which parameters are able to predict whether to continue the ART path after COS phase to obtain biochemical pregnancy/clinical pregnancy/live birth? |
| Second step | Coding of predictors | Variables are divided in baseline characteristics and ART variables. All categorical variables have been transformed in binary factors |
| Third step | Model specification | Step 1: logistic linear regression analysis with stepwise selection method Step 2: decision trees analysis applying Wald statistics |
| Fourth step | Model estimation | Step 1: Spearman's Rho Step 2: Chi-squared |
| Fifth step | Model performance | Not applicable in this clinical context |
| Sixth step | Model validity | A random sample including 50% of the cohort was selected for model development and the remaining 50% for the validation. Then, the validation was confirmed by cross-validation resampling |
| Seventh step | Model presentation | Decision trees construction for each strong ART outcomes |

ART, assisted reproductive techniques; COS, controlled ovarian stimulation

converted into ordinal predictors before the merging step. The primary advantage CHAID decision tree analysis is the large number of variables potentially usable in the segmentation process. In these analyses, nodes were created considering those variables acting before the ovulation induction, such as female and male ages, BMI, smoking habit, infertility causes, gonadotropin drug used, starting gonadotropin dose, days of stimulation, total gonadotropin units used, and OF17. The analysis was performed using the 50% of the casuistry, randomly selected, to test the tree, and the remnant 50% to validate the result. Percentages reported within each node of the decision tree generated did not report the occurrence of the endpoint evaluated but the accuracy of the classification performed by each node. Since this approach could be suboptimal for internal validation, we confirmed these results applying a further cross-validation resampling [22]. Moreover, in applying the CHAID algorithm, the rule of thumb (or stopping rule) for the growth of the tree had a key role. Thus, we considered a minimum sample size of 50 cases for the terminal nodes (final segments). As such, we assure the assumption of normality for an ANOVA procedure to compare the means of a continuous variable of interest for each segment. In addition, if the variable of interest is a categorical one, we reach a reasonable sampling size to apply a multinomial logistic regression.

Finally, since the dataset included couples treated one time and couples in which ART was applied more than two times, the analyses were repeated considering the couples treated only one time as a single group.

Statistical analysis was performed using the “Statistical Package for the Social Sciences” software for Windows (version 26.0; SPSS Inc., Chicago, IL, USA). For all comparisons, $p < 0.05$ was considered statistically significant.

Results

The final database included 12,275 ART cycles, consisting of 7826 ICSI (63.8%) and 4449 IVF (36.2%) procedures. The 87.5% of the entire cohort (10,375 couples) were treated for primary couple infertility. Table 2 summarizes main baseline cohort characteristics of couples enrolled. Table 3 shows the ART variables and outcomes (as distinguished above) collected for each ART cycle included in the analyses.

Comparing couples who obtained a biochemical pregnancy to those who did not, the pregnant couples showed lower female and male ages ($p < 0.001$ and $p = 0.002$, respectively), higher sperm morphology ($p = 0.044$), OF17 ($p < 0.001$), total retrieved oocytes ($p < 0.001$), injected/inseminated oocytes ($p < 0.001$), fertilized oocytes ($p < 0.001$), total embryos ($p < 0.001$), transferred embryos ($p < 0.001$), and fertilization rate ($p < 0.001$) (Table 4). Similarly, couples who achieved a clinical pregnancy showed lower female and male ages ($p = 0.002$ and $p < 0.001$, respectively), higher sperm morphology ($p = 0.006$), number of ovarian follicles higher than 17 mm ($p < 0.001$), total retrieved oocytes ($p = 0.025$), injected/inseminated oocytes ($p = 0.004$), oocytes fertilized ($p < 0.001$), total embryos ($p < 0.001$), fertilized embryos ($p < 0.001$), and fertilization rate ($p < 0.001$) (Table 5). On the contrary, couples who achieved a live birth showed fewer statistically significant differences compared to couples who did not, as expected. In particular, couples who achieved a live birth showed higher OF17 number ($p < 0.001$), total retrieved oocytes ($p < 0.001$), and total embryos ($p < 0.001$) (Table 6).

Linear logistic regression analysis detected three ART variables able to influence the biochemical pregnancy

Table 2 Cohort baseline characteristics

| Couples' characteristics | Mean ± standard deviation |
|---|---------------------------|
| Number of cycles (<i>n</i>) | 12,275 |
| Number of couples performing two cycles | 3122 |
| Number of couples performing more than two cycles | 572 |
| Female age (years) | 36.9 ± 4.5 |
| Female body mass index (kg/m ²) | 23.9 ± 4.8 |
| Male age (years) | 40.2 ± 5.5 |
| Male body mass index (kg/m ²) | 23.8 ± 7.5 |
| Infertility causes | |
| <i>Pelvic n</i> (%) | 862 (7.0%) |
| <i>Endometriosis n</i> (%) | 1085 (8.8%) |
| <i>Tubal n</i> (%) | 1315 (10.7%) |
| <i>Idiopathic n</i> (%) | 3263 (26.6%) |
| <i>Advanced maternal age n</i> (%) | 2308 (18.8%) |
| <i>Male factors n</i> (%) | 5174 (42.2%) |
| Previous pregnancies | |
| 1 <i>n</i> (%) | 2961 (24.1%) |
| 2 <i>n</i> (%) | 1165 (9.5%) |
| 3 <i>n</i> (%) | 446 (3.6%) |
| 4 <i>n</i> (%) | 181 (1.5%) |
| > 5 <i>n</i> (%) | 76 (0.6%) |
| Previous miscarriages | |
| 1 <i>n</i> (%) | 2588 (21.1%) |
| 2 <i>n</i> (%) | 917 (7.5%) |
| 3 <i>n</i> (%) | 282 (2.3%) |
| 4 <i>n</i> (%) | 111 (0.9%) |
| > 5 <i>n</i> (%) | 48 (0.3%) |
| Previous pre-term delivery | |
| 1 <i>n</i> (%) | 114 (0.9%) |
| 2 <i>n</i> (%) | 6 (0.1%) |
| Previous delivery | |
| 1 <i>n</i> (%) | 1386 (11.3%) |
| 2 <i>n</i> (%) | 121 (1.0%) |
| 3 <i>n</i> (%) | 19 (0.2%) |
| 4 <i>n</i> (%) | 13 (0.1%) |
| Smoking habit | |
| <i>Female n</i> (%) | 994 (8.1%) |
| <i>Male n</i> (%) | 1252 (10.2%) |

rate, such as OF17 number ($B = 0.336$; Wald coefficient = 887.1; $p < 0.001$; OR, 1.40; CI95%, 1.37–1.43), the number of total embryos ($B = 0.132$; Wald coefficient = 58.0; $p < 0.001$; OR, 1.14; CI95%, 1.10–1.18), and the fertilization rate ($B = 0.982$; Wald coefficient = 91.9; $p < 0.001$; OR, 2.67; CI95%, 2.18–3.26). Similar results were obtained considering the clinical pregnancy rate as dependent variable. However, the latter was significantly influenced by the OF17 number ($B = 0.494$;

Table 3 Assisted reproductive technique (ART) variables and outcomes. Data are expressed as mean ± standard deviation

| ART variables | |
|---|-----------------|
| Days of stimulation (<i>n</i>) | 13.1 ± 10.2 |
| Total FSH units (<i>n</i>) | 2879.8 ± 2011.3 |
| Estradiol (pg/mL) | 1562.9 ± 1004.4 |
| OF17 (<i>n</i>) | 4.3 ± 2.5 |
| Semen volume (mL) | 3.0 ± 1.9 |
| Sperm concentration (million/mL) | 35.0 ± 37.2 |
| Sperm motility (%) | 27.7 ± 18.6 |
| Sperm morphology (%) | 3.8 ± 3.6 |
| Total oocytes retrieved (<i>n</i>) | 5.5 ± 3.6 |
| Mature oocytes (<i>n</i>) | 5.5 ± 4.2 |
| Oocytes injected/inseminated (<i>n</i>) | 4.5 ± 3.2 |
| Fertilized oocytes (<i>n</i>) | 3.6 ± 2.3 |
| Fertilization rate (%) | 60.0 ± 34 |
| Total embryos (<i>n</i>) | 2.7 ± 2.3 |
| Transferred embryos (<i>n</i>) | 1.6 ± 1.3 |
| ART outcomes | |
| Pregnancy (biochemical) <i>n</i> (%) | 2817 (22.9) |
| Pregnancy (clinic) <i>n</i> (%) | 2442 (19.9) |
| Live birth <i>n</i> (%) | 2212 (18.0) |

ART, assisted reproductive techniques; FSH, follicle-stimulating hormone; OF17, ovarian follicles > 17 mm detected at ultrasound before pick-up

Wald coefficient = 281.0; $p < 0.001$; OR, 1.64; CI95%, 1.55–1.73), and the total number of embryos formed ($B = 0.140$; Wald coefficient = 14.2; $p < 0.001$; OR, 1.15; CI95%, 1.07–1.24), but not by the fertilization rate. Finally, live birth rate was not significantly related to any ART variable in logistic regression analysis.

Thus, three ART variables are related to ART outcomes, such as ovarian follicle number at US, total embryos, and fertilization rate. These parameters were set as dependent variables in correlation analyses with all parameters that chronologically precede them. The OF17 number was inversely related to female age ($p < 0.001$), BMI ($p < 0.001$), and total FSH dose used for COS ($p < 0.001$) (Table 3). On the contrary, neither the FSH starting dose, nor the length of COS significantly influenced the OF17 number (Table 7). Similarly, the number of embryos obtained was inversely related to the female age ($p < 0.001$) and directly related to the OF17 number ($p < 0.001$), semen volume ($p < 0.001$), sperm concentration ($p < 0.001$), sperm motility ($p < 0.001$), sperm morphology ($p < 0.001$), total retrieved oocytes ($p < 0.001$), mature oocytes ($p < 0.001$), injected/inseminated oocytes ($p < 0.001$), and fertilized oocytes ($p < 0.001$) (Table 7). Finally, the fertilization rate was significantly directly related to OF17 ($p < 0.001$), sperm concentration ($p < 0.001$), sperm motility ($p < 0.001$), sperm morphology

Table 4 Comparison between couples who achieved a biochemical pregnancy compared to couples who did not. Data are expressed as mean \pm standard deviation. Bold values express statistically significant differences

| Biochemical pregnancy | | | |
|---|--|--|-------------------|
| ART variables | Biochemically pregnant couples (<i>n</i> = 2817) | Biochemically not pregnant couples (<i>n</i> = 9458) | <i>p</i> -value |
| Female age (years) | 36.4 \pm 4.1 | 37.1 \pm 4.1 | < 0.001 |
| Female body mass index (kg/m ²) | 24.8 \pm 1.9 | 25.0 \pm 1.3 | 0.621 |
| Male age (years) | 39.9 \pm 5.3 | 40.3 \pm 5.5 | 0.002 |
| Days of stimulation (<i>n</i>) | 13.6 \pm 10.9 | 12.8 \pm 4.8 | 0.543 |
| Total FSH units (<i>n</i>) | 2433.8 \pm 1677.9 | 2983.8 \pm 2067.7 | 0.487 |
| Semen volume (mL) | 2.9 \pm 2.3 | 2.8 \pm 1.8 | 0.219 |
| Sperm concentration (million/mL) | 33.7 \pm 33.8 | 34.0 \pm 37.1 | 0.686 |
| Sperm motility (%) | 28.1 \pm 18.1 | 27.9 \pm 18.1 | 0.708 |
| Sperm morphology (%) | 3.9 \pm 3.7 | 3.7 \pm 3.5 | 0.044 |
| OF17 (<i>n</i>) | 5.9 \pm 2.4 | 3.9 \pm 2.3 | < 0.001 |
| Total oocytes retrieved (<i>n</i>) | 6.1 \pm 3.0 | 5.3 \pm 3.7 | < 0.001 |
| Mature oocytes (<i>n</i>) | 5.6 \pm 4.5 | 5.5 \pm 4.2 | < 0.001 |
| Oocytes injected/inseminated (<i>n</i>) | 5.2 \pm 2.8 | 4.3 \pm 3.2 | < 0.001 |
| Fertilized oocytes (<i>n</i>) | 3.5 \pm 1.9 | 2.4 \pm 2.3 | < 0.001 |
| Fertilization rate (%) | 70.0 \pm 29 | 57.0 \pm 34 | < 0.001 |
| Total embryos (<i>n</i>) | 3.6 \pm 2.0 | 2.4 \pm 2.3 | < 0.001 |
| Transferred embryos (<i>n</i>) | 2.4 \pm 1.0 | 1.4 \pm 1.2 | < 0.001 |

ART, assisted reproductive techniques; FSH, follicle-stimulating hormone; OF17, ovarian follicles > 17 mm detected at ultrasound before pick-up

Table 5 Comparison between couples who achieved a clinical pregnancy compared to couples who did not. Data are expressed as mean \pm standard deviation. Bold values express statistically significant differences

| Clinical pregnancy | | | |
|---|---|---|-------------------|
| ART variables | Clinically pregnant couples (<i>n</i> = 2442) | Clinically not pregnant couples (<i>n</i> = 9883) | <i>p</i> -value |
| Female age (years) | 36.2 \pm 4.1 | 36.9 \pm 4.2 | 0.002 |
| Female body mass index (kg/m ²) | 24.7 \pm 1.9 | 25.1 \pm 1.4 | 0.437 |
| Male age (years) | 39.6 \pm 5.1 | 40.6 \pm 5.7 | < 0.001 |
| Days of stimulation (<i>n</i>) | 13.7 \pm 13.9 | 13.2 \pm 5.2 | 0.581 |
| Total FSH units (<i>n</i>) | 2328.4 \pm 1406.8 | 2743.6 \pm 2319.1 | 0.145 |
| Semen volume (mL) | 2.9 \pm 2.5 | 2.9 \pm 1.7 | 0.805 |
| Sperm concentration (million/mL) | 34.3 \pm 34.2 | 31.6 \pm 31.8 | 0.107 |
| Sperm motility (%) | 28.5 \pm 18.4 | 26.9 \pm 17.1 | 0.089 |
| Sperm morphology (%) | 3.9 \pm 3.2 | 3.7 \pm 3.3 | 0.006 |
| OF17 (<i>n</i>) | 6.4 \pm 2.3 | 4.2 \pm 2.2 | < 0.001 |
| Total oocytes retrieved (<i>n</i>) | 6.2 \pm 3.0 | 5.9 \pm 3.1 | 0.025 |
| Mature oocytes (<i>n</i>) | 5.7 \pm 4.7 | 5.4 \pm 4.3 | 0.004 |
| Oocytes injected/inseminated (<i>n</i>) | 5.3 \pm 2.8 | 5.0 \pm 2.9 | < 0.001 |
| Fertilized oocytes (<i>n</i>) | 3.5 \pm 1.9 | 3.3 \pm 2.0 | < 0.001 |
| Fertilization rate (%) | 71.0 \pm 28 | 67.0 \pm 39 | < 0.001 |
| Total embryos (<i>n</i>) | 3.7 \pm 2.0 | 3.3 \pm 2.0 | < 0.001 |
| Transferred embryos (<i>n</i>) | 2.4 \pm 1.0 | 2.2 \pm 1.0 | < 0.001 |

ART, assisted reproductive techniques; FSH, follicle-stimulating hormone; OF17, ovarian follicles > 17 mm detected at ultrasound before pick-up

Table 6 Comparison between couples who achieved a live birth compared to couples who did not. Data are expressed as mean \pm standard deviation. Bold values express statistically significant differences

| ART variables | Live birth | | <i>p</i> -value |
|---|--|---|-------------------|
| | Couples with live birth (<i>n</i> = 2212) | Couples without live birth (<i>n</i> = 10,063) | |
| Female age (years) | 36.7 \pm 4.7 | 36.9 \pm 4.8 | 0.722 |
| Female body mass index (kg/m ²) | 24.9 \pm 1.8 | 25.0 \pm 1.6 | 0.498 |
| Male age (years) | 39.6 \pm 5.1 | 40.6 \pm 5.7 | 0.080 |
| Days of stimulation (<i>n</i>) | 13.9 \pm 23.9 | 11.7 \pm 7.2 | 0.555 |
| Total FSH units (<i>n</i>) | 2321.7 \pm 1402.3 | 2230.1 \pm 1221.6 | 0.217 |
| Semen volume (mL) | 2.9 \pm 2.6 | 2.6 \pm 1.9 | 0.289 |
| Sperm concentration (million/mL) | 33.8 \pm 34.8 | 32.6 \pm 33.1 | 0.282 |
| Sperm motility (%) | 28.3 \pm 18.5 | 27.1 \pm 18.4 | 0.159 |
| Sperm morphology (%) | 3.9 \pm 3.7 | 3.4 \pm 2.6 | 0.054 |
| OF17 (<i>n</i>) | 6.5 \pm 2.3 | 5.5 \pm 1.9 | < 0.001 |
| Total oocytes retrieved (<i>n</i>) | 6.2 \pm 3.0 | 5.8 \pm 2.9 | < 0.001 |
| Mature oocytes (<i>n</i>) | 5.8 \pm 5.0 | 5.6 \pm 4.5 | 0.236 |
| Oocytes injected/inseminated (<i>n</i>) | 5.3 \pm 2.8 | 5.0 \pm 2.9 | 0.126 |
| Fertilized oocytes (<i>n</i>) | 3.5 \pm 1.9 | 3.3 \pm 2.1 | 0.076 |
| Fertilization rate (%) | 71.0 \pm 31 | 68.0 \pm 32 | 0.093 |
| Total embryos (<i>n</i>) | 3.7 \pm 1.9 | 3.2 \pm 2.1 | < 0.001 |
| Transferred embryos (<i>n</i>) | 2.4 \pm 1.0 | 2.4 \pm 0.9 | 0.909 |

ART, assisted reproductive techniques; FSH, follicle-stimulating hormone; OF17, ovarian follicles > 17 mm detected at ultrasound before pick-up

($p < 0.001$), total retrieved oocytes ($p < 0.001$), injected/inseminated oocytes ($p < 0.001$), fertilized oocytes ($p < 0.001$), total embryos ($p < 0.001$), transferred embryos ($p < 0.001$) (Table 7).

The three ART variables that significantly influenced pregnancy and live birth rates were then used as dependent variables in multivariate linear regression analyses, setting other factors, and baseline characteristics as independent variables. No statistical models were generated by multivariate analyses able to correlate the three ART variables with other parameters. These results suggest that the three parameters are not statistically influenced by any of the other ART parameters considered.

Decision trees analysis

The first decision tree was created using biochemical pregnancy as dependent variable. The statistical power of the analysis was 80.8% for the training (relative risk = 0.192, standard error 0.005) and 81.4% for the validation analysis (relative risk = 0.186, standard error 0.005). Seven predictive nodes have been identified (Fig. 1). The first five nodes classified biochemical pregnancies according to thresholds for OF17 (Fig. 1). The result of the decision tree suggests that at least two thresholds of OF17 could be suggested. When two or less follicles have been recognized at US before pick-up a biochemical pregnancy is virtually unachievable, while when more than 7 follicles are observed, there is the highest

probability to achieve a biochemical pregnancy. Moreover, although between three and seven follicles at US, no clear distinction between pregnancy and no pregnancy could be achieved, when the follicle numbers at US was three, sperm motility entered the model (nodes 6 and 7), detecting a predicting threshold of 34.0% of progressive sperm motility (Fig. 1).

The second decision tree was performed on clinical pregnancy, reaching a statistical power of 85.4% for the training (relative risk = 0.146, standard error 0.005) and 86.8% for the validation analysis (relative risk = 0.138, standard error 0.004). Ten predictive nodes were identified (Fig. 2), in which the OF17 number remained in the first six nodes (Fig. 2). As for biochemical pregnancy, this analysis confirmed two thresholds regarding the OF17 number, i.e., equal or below 2 and above 7 follicles. In this analysis, between 5 and 7 follicles detected at US, no clear distinction between pregnancy and no pregnancy could be detected. On the contrary, when the number of follicles was 3 or 4/5, the female age entered the model with two different thresholds, respectively (37.2 years, nodes 7 and 8, and 40.1 years, nodes 9 and 10) (Fig. 2).

Finally, the third decision tree analysis was made for live birth rate. In this last model, the OF17 represented the first five nodes, with a statistical power of 86.6% for the training (relative risk = 0.134, standard error 0.004) and 87.2% for the validation analysis (relative risk = 0.128, standard error 0.004) (Fig. 3). This result suggests that

Table 7 Linear correlation analysis between parameters which predicted assisted reproduction techniques (ART) outcomes and all parameters preceding the variable itself. Bold values express statistically significant differences

| ART variables | | OF17 | Total embryos | Fertilization rate |
|------------------------------|-----------------|------------------|------------------|--------------------|
| Female age | Spearman's Rho | −0.100 | −0.495 | 0.021 |
| | <i>p</i> -value | <0.001 | <0.001 | 0.064 |
| Male age | Spearman's Rho | - | 0.004 | 0.001 |
| | <i>p</i> -value | - | 0.523 | 0.943 |
| Female BMI | Spearman's Rho | −0.090 | 0.002 | −0.018 |
| | <i>p</i> -value | <0.001 | 0.243 | 0.054 |
| FSH starting dose | Spearman's Rho | −0.036 | 0.087 | 0.038 |
| | <i>p</i> -value | 0.121 | 0.231 | 0.381 |
| Days of stimulation | Spearman's Rho | −0.003 | −0.001 | 0.110 |
| | <i>p</i> -value | 0.768 | 0.589 | 0.232 |
| Total FSH units | Spearman's Rho | −0.224 | 0.058 | −0.044 |
| | <i>p</i> -value | <0.001 | 0.081 | 0.075 |
| OF17 | Spearman's Rho | - | 0.281 | 0.312 |
| | <i>p</i> -value | - | <0.001 | <0.001 |
| Semen volume | Spearman's Rho | - | 0.007 | 0.025 |
| | <i>p</i> -value | - | 0.435 | 0.117 |
| Sperm concentration | Spearman's Rho | - | 0.089 | 0.132 |
| | <i>p</i> -value | - | <0.001 | <0.001 |
| Sperm motility | Spearman's Rho | - | 0.058 | 0.144 |
| | <i>p</i> -value | - | <0.001 | <0.001 |
| Sperm morphology | Spearman's Rho | - | 0.051 | 0.187 |
| | <i>p</i> -value | - | <0.001 | <0.001 |
| Total oocytes retrieved | Spearman's Rho | - | 0.634 | 0.182 |
| | <i>p</i> -value | - | <0.001 | <0.001 |
| Mature oocytes | Spearman's Rho | - | 0.677 | 0.043 |
| | <i>p</i> -value | - | <0.001 | 0.153 |
| Oocytes injected/inseminated | Spearman's Rho | - | 0.732 | 0.290 |
| | <i>p</i> -value | - | <0.001 | <0.001 |
| Fertilized oocytes | Spearman's Rho | - | 0.932 | 0.543 |
| | <i>p</i> -value | - | <0.001 | <0.001 |
| Total embryos | Spearman's Rho | - | - | 0.458 |
| | <i>p</i> -value | - | - | <0.001 |
| Transferred embryos | Spearman's Rho | - | - | 0.363 |
| | <i>p</i> -value | - | - | <0.001 |

ART, assisted reproductive techniques; BMI, body mass index; FSH, follicle-stimulating hormone; OF17, ovarian follicles > 17 mm detected at ultrasound before pick-up

OF17 is the first parameter able to classify pregnant and not pregnant women (first node). However, only the threshold of 2 follicles is confirmed as a predictor of lack ART success. On the contrary, other nodes did not clearly classify live birth rate. In this analysis, if the OF17 number was 3 or > 7, female age entered into the model (second node) with the thresholds of 37.2 (nodes 6 and 7) and 35.6 years (nodes 8 and 9), respectively (Fig. 3).

Decision tree analyses for each strong ART outcome were repeated considering only those couples in which the ART was applied only one time (i.e., 7952 couples).

The predictive accuracy of all three trees remained also considering only this subgroup (data not shown).

Male contribution

Considering the role of sperm motility in the biochemical pregnancy decision tree, the casuistry was divided according to the threshold suggested (i.e., 34%). Fisher exact test was performed to compare biochemical and clinical pregnancy between the two groups created, considering IVF and ICSI separately. Indeed, we previously demonstrated

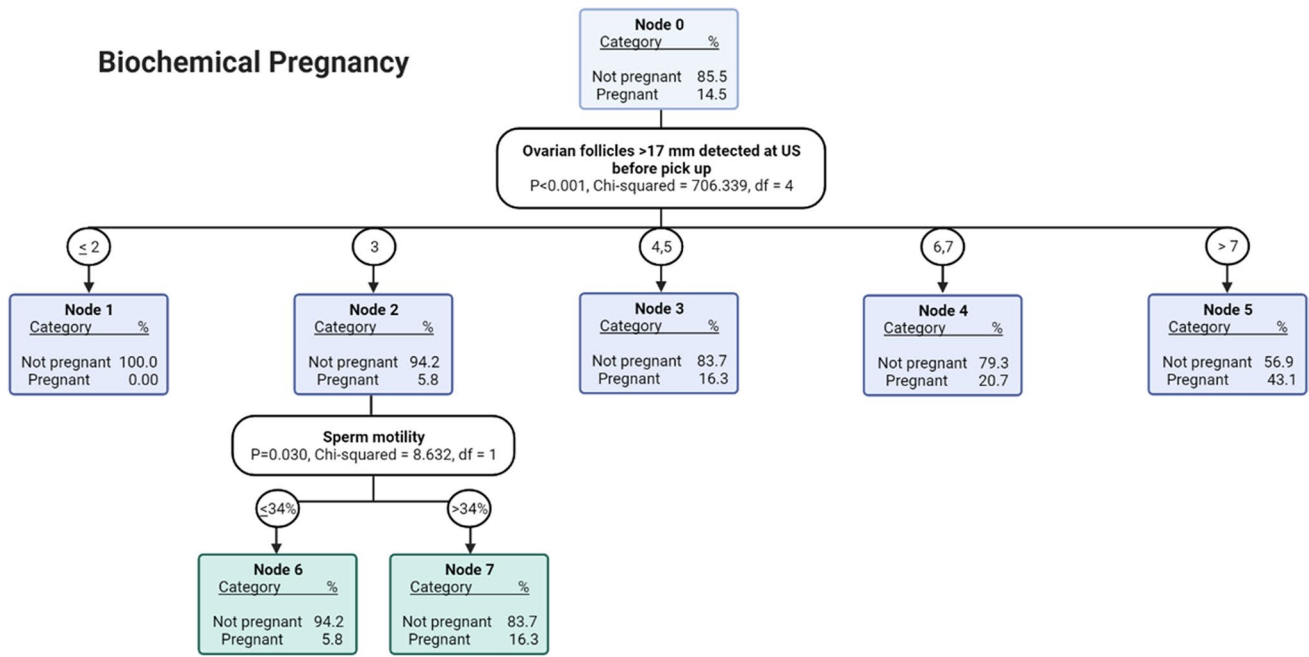


Fig. 1 Decision tree performed using biochemical pregnancy as dependent variable. Only the validation step of the analysis is represented. Percentages reported for each node express the predictive accuracy of the node. Df, degrees of freedom; US, ultrasound

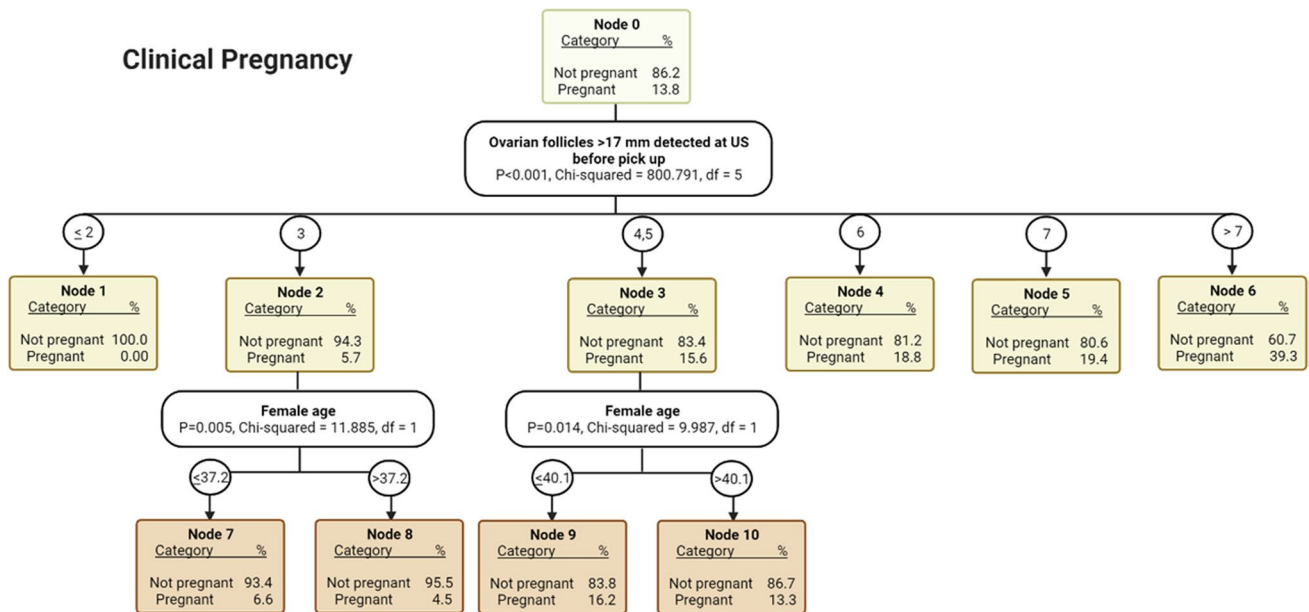


Fig. 2 Decision tree performed using clinical pregnancy as dependent variable. Only the validation step of the analysis is represented. Percentages reported for each node express the predictive accuracy of the node. Df, degrees of freedom; US, ultrasound

that sperm motility could have a predictive role in IVF, rather than ICSI cycles (Villani et al. 2021, submitted). As a confirm, both biochemical (75.3 versus 24.7%, $p < 0.001$) and clinical pregnancy (74.8 versus 25.2%, $p < 0.001$)

rates were significantly higher in couples in which the man showed sperm motility higher than 34% compared to the others only in IVF cycles. On the contrary, no differences in biochemical (44.1 versus 55.9%, $p = 0.189$) and

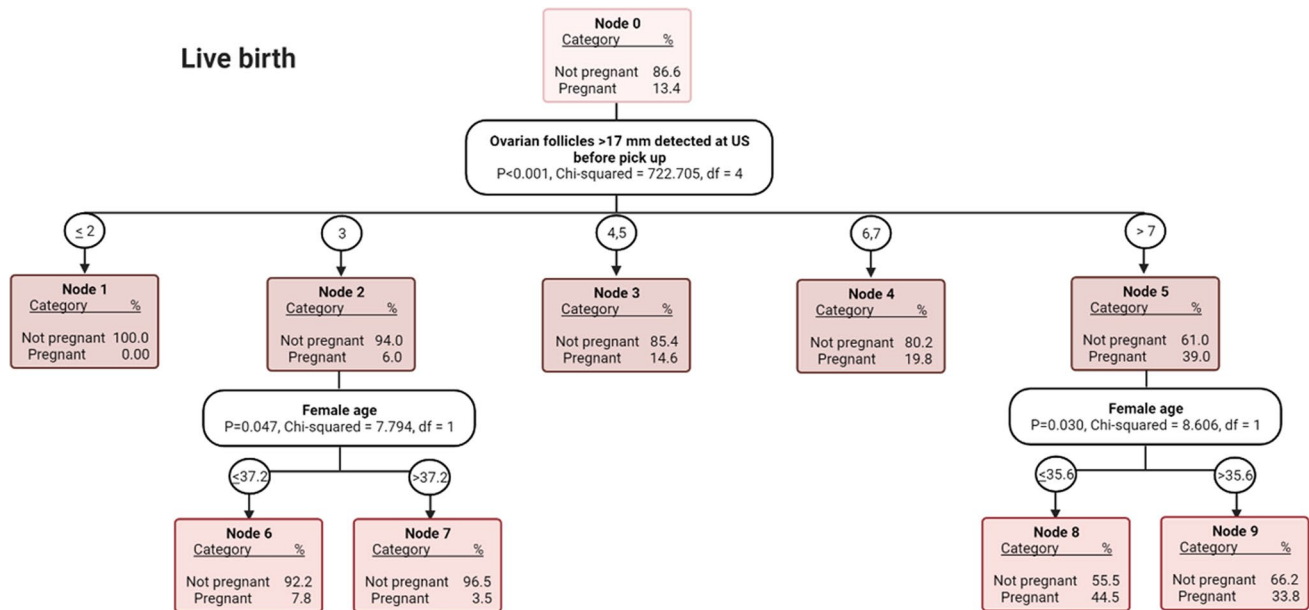


Fig. 3 Decision tree performed using live birth as dependent variable. Only the validation step of the analysis is represented. Percentages reported for each node express the predictive accuracy of the node. Df, degrees of freedom; US, ultrasound

clinical (44.8 versus 55.2%, $p = 0.303$) pregnancy rates were detected between the two groups created on sperm motility in case of ICSI.

Discussion

Here, we applied a systematic seven-steps approach to generate a predicting ART success model [20], detecting three sensitive milestones of the decision-making process in which the clinician is routinely involved. In particular, OF17 number, female age, and sperm motility could be used to evaluate whether the ART path should be continued before pick-up. Several mathematical and statistical models have been proposed so far to predict ART success. However, an overall limited predictive accuracy and clinical utility emerged, due to several shortcomings and to a probably incorrect view to the problem [23]. Indeed, these previous works were finalized to identify predicting markers of overall ART success, not asking the question of how these factors could influence the decision-making process. Here, on the other hand, we have changed the point of view to the question, first selecting the ART path sensitive points, where a predictor could advise the doctor either to suspend the treatment or to change the approach. Two ART sensitive milestones, in which the clinician could decide to stop the process, could be the pick-up and the transfer time. Here, we demonstrate that the decision to continue the ART path to pick-up could be guided by three specific factors, applying a decision tree analyses.

Logistic regression analyses confirmed the relationship between strong ART outcomes and those variables detected before pick-up. Interestingly, these connections appeared only when pregnancy rates were considered, suggesting that the classical statistical approach is not able to overcome the higher number of biases influencing live birth rates. In the biochemical pregnancy decision tree, alongside to OF17, sperm motility entered the model introducing the threshold of 34%. This result suggests that a male parameter represents a crucial point in terms of prediction of pregnancy obtainment with a cut-off near to what proposed by the WHO manual, i.e., 32% [24, 25]. Although the relevance of male contribution in human reproduction seems obvious, most studies aimed at predicting ART success relegated the male factor to a secondary role, evaluating only the female partner (Villani MT et al., 2021, submitted). Together with the male partner, two female-related parameters emerged as the strongest and most clinically relevant key points derived from our decision trees, i.e., OF17 and female age. In particular, when less or equal than two follicles have been identified by US after COS, the chance of pregnancy obtainment is virtually zero. Similarly, in case of three OF17, the chance of conceiving remained below 6% for all three developed trees. On the other hand, increasing the OF17 number coherently, the pregnancy probabilities raised, however without identifying a clear threshold beyond which the virtual certainty of pregnancy obtainment is reached. When more than seven follicles have been identified, the pregnancy probabilities are the highest possible. Moreover, our trees showed that, for intermediate OF17 results, the female age could guide

the clinician decision. In other words, our results suggest that when OF17 is lower than three the ART path should be stopped and the COS re-scheduled. When the OF17 is higher than 7, the ART path should be followed. When the OF17 is between 3 and 6, other parameters should be considered. In particular, in this setting, when the female age is high or when the sperm motility is low, the chance of ART success significantly decreases, and a COS re-schedule should be considered.

Recently, a meta-analysis evaluated the methodological quality and the performance of all existing ART predictive models, to recommend the most accurate approach at predicting chances of parenting after ART procedures, helping couples in managing their expectations [23]. The first interesting result of this meta-analysis is the high heterogeneity of statistical approach to the topic. Indeed, considering 35 predictive models from 33 manuscripts, logistic regression analyses were applied in 91% of studies and time to event modelling in 9% [23, 26, 27]. A combined statistical approach was not applied in all these analyses. The second relevant issue raised by this meta-analysis was the sample size considered in the 35 trials included. Indeed, only four works (11.4%) had sample sizes large enough to support the development and the validation of their models [18, 28–30], while the remaining analyzed cohorts of patients was insufficient to obtain reliable predictors. Here, we considered more than twelve thousand fresh cycles, applying a complex statistical model in which the linear logistic regression analysis was combined to a decision tree classification approach. Moreover, the vast majority of previous predictive models shares the limitation of not being processed in a systematic way, avoiding the recommended methodological development steps and consequently limiting their reliability [20, 31, 32]. In our work, we applied a systematic seven-steps approach to predictive model development [20] on a consistent single center casuistry of fresh ART cycles. In this systematic approach, the first step provides the problem definition and data inspection. Evaluating what is already known in the literature, the problem has been often faced looking for those parameters able to predict the final success, not thinking about how these can then be really applied to clinical practice. Thus, we changed the point of view on the problem, focusing the analysis on the search for those parameters able to tell the clinician when to continue the ART path after the COS phase, i.e., whether to proceed or not with the pick-up. Thus, our model could help to objectively predict a priori the potentiality in terms of pregnancy/live birth after ovarian stimulation. In particular, our model showed that three variables (OF17, female age, and sperm motility) could help the clinician to decide when and whether to continue the ART path after COS, and, consequently, when re-schedule a new ART cycle. This change of view could be potentially extremely relevant in ART management, to

avoid loss of time and money stopping the ART cycle at an early stage when success chances result negligible. As a consequence, a new ART cycle could be planned to modify the COS phase in order to improve the OF17 number and to proceed to pick-up with higher probability of success. Indeed, the COS phase is heterogeneously managed comparing different assisted fertilization centers given the absence of evidence-based protocols, but crucial for ART success [13, 33]. Apart from tangible clinic and economic advantages, possible repercussions on the psychological health of the infertile couple have to be taken into account. Indeed, it is well established in the scientific literature and clearly evident in clinical practice that ART procedures are accompanied by a significant emotional burden experienced by both partners [34]. Since psychological consequences could be even more burdensome in case of ART treatment failures and with the consequent need to schedule new cycles [34], a precocious suspension of the ART cycle followed by a new treatment schedule could improve the couples' psychological health.

Our study fits well into this research line [35], applying a validated statistical method and enriching it with the analysis of decision trees. A decision tree is a tree-like model commonly used as a supportive tool in operation research and decision analyses, such as in economy and marketing settings. This approach, aimed at identifying the strategy most likely to reach a goal, is simple to understand and interpret and could be combined with other decision techniques, such as logistic regression analyses. For these reasons, the decision tree model has been applied in several medicine branches, such as gastroenterology [36, 37], breast oncological surgery [38], cardiology [39], orthopedic surgery [40], and even to diagnose SARS-CoV-2 infection [41]. Specifically in the ART setting, the decision tree model has been previously applied mainly in cost-effectiveness analyses, for instance to identify the most cost-effective ovarian stimulation drug for intra-uterine insemination (IUI) [42]; to evaluate the clinical utility for preimplantation genetic assessment for aneuploidy after IVF in the USA [43], and in Germany [44]; to highlight anti-Müllerian hormone (AMH) serum levels as informative for stimulation dose management for optimizing blastocyst development [45]; and to identify the most cost-effective policy in terms of ART success in case of female age below 38 years comparing expectant management, IUI with ovarian stimulation and IVF [46]. Moreover, a decision tree was applied to develop a model combining AMH, antral follicle counts, FSH basal levels and female age to obtain the true ovarian reserve [47], and to compare GnRH-agonist long protocol to GnRH-antagonist protocol in IVF, highlighting that GnRH-antagonist introduces an economic advantage in case of fresh embryos, while the GnRH-agonist long protocol is preferable considering the cumulative pregnancy rate using both fresh and frozen

embryos [48]. Here, we applied for the first time this statistical approach to new clinical question that should be increasingly relevant in ART practice.

The main strengths of our study are represented by the large sample size and the systematic approach to the predictive model development. However, several limitations should be considered. First, the retrospective collection of real-world data is biased by a high rate of missing data, possibly impacting the reliability of the results. However, we included only those cycles in which all ART variables were available, at the cost of reducing the sample size. In addition, during the long interval of data collection (i.e., 1998–2020), ART technologies evolved, as well as the regulatory rules for ART access and the characteristics of couples recurring to ART procedures. This data heterogeneity over the years could mitigate the reliability of our results. Moreover, we decided to develop our models excluding frozen embryos to avoid possible confounding factors, whereby obtained results are reliable only for fresh ART cycles. However, the large time-frame interval of observations, together with the use of only fresh cycles could be the reason for the low overall pregnancy rates detected in the cohort. Finally, our results come from a fairly young population, limiting their actual application to an older cohort. The accuracy of our model is 100% when less than three OF17 were detected and reaches the 43% when more than three OF17 were identified. However, this is true in our casuistry and further studies should confirm this accuracy.

In conclusion, we identified three decision trees helping the clinician to decide whether or not to perform oocytes pick-up, continuing the ongoing ART path. In these mathematical models, three predictors of ART success at a very early stage emerged, such as OF17 number, sperm motility, and female age. Although the female age constitutes a non-modifiable factor, the increase of OF17 and sperm motility should be pursued by clinicians to improve the chances of ART success.

Author contribution MTV, DM, BM, AN, FI, GBLS, and LA were involved in acquisition of data and revised the article critically for important intellectual content; MS revised the article critically for important intellectual content; MTV, DM, GS, CF, and DS conceived and designed the study, analyzed and interpreted data, and drafted the article. All authors approved the final version of the manuscript to be submitted.

Data availability Data supporting the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Declarations

Ethics approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study protocol was approved by the local Ethics Committee “Area Vasta Nord of the Emilia Romagna” (protocol number AOU 0024637/19).

Consent to participate/consent for publication Considering the retrospective study design, informed consent was waived by the Ethics Committee, and data were collected anonymously.

Conflict of interest The authors declare no competing interests.

References

1. Vander Borgh M, Wyns C. Fertility and infertility: definition and epidemiology. *Clin Biochem.* 2018;62:2–10.
2. Saadati N, Nikbakht R, Sattari A, Amininezhad FS. Spontaneous pregnancies and determinant factors in infertility: a cross-sectional study. *Int J Reprod Biomed.* 2020;18(10):905–10.
3. De Geyter C. More than 8 million babies born from IVF since the world's first in 1978: European IVF pregnancy rates now steady at around 36 percent, according to ESHRE monitoring. *European Society of Human Reproduction and Embryology*; 2018.
4. Domar AD. Impact of psychological factors on dropout rates in insured infertility patients. *Fertil Steril.* 2004;81(2):271–3.
5. Moragianni VA, Penzias AS. Cumulative live-birth rates after assisted reproductive technology. *Curr Opin Obstet Gynecol.* 2010;22(3):189–92.
6. Yang J, Zhang X, Ding X, Wang Y, Huang G, Ye H. Cumulative live birth rates between GnRH-agonist long and GnRH-antagonist protocol in one ART cycle when all embryos transferred: real-world data of 18,853 women from China. *Reprod Biol Endocrinol.* 2021;19(1):124.
7. Banker M, Dyer S, Chambers GM, Ishihara O, Kupka M, de Mouzon J, Zegers-Hochschild F, Adamson GD. International Committee for Monitoring Assisted Reproductive Technologies (ICMART): world report on assisted reproductive technologies, 2013. *Fertil Steril.* 2021;116(3):741–56.
8. van der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Bossuyt PM, Hompes PG, van der Veen F, Mol BW. Do clinical prediction models improve concordance of treatment decisions in reproductive medicine? *BJOG.* 2006;113(7):825–31.
9. Khader A, Lloyd SM, McConnachie A, Fleming R, Grisendi V, La Marca A, Nelson SM. External validation of anti-Müllerian hormone based prediction of live birth in assisted conception. *J Ovarian Res.* 2013;6(1):3.
10. Wyns C, Bergh C, Calhaz-Jorge C, De Geyter C, Kupka MS, Motrenko T, Rugescu I, Smeenk J, Tandler-Schneider A, Vidakovic S, et al. ART in Europe, 2016: results generated from European registries by ESHRE. *Hum Reprod Open.* 2020;2020(3):hoaa032.
11. Yakin K, Oktem O, Balaban B, Urman B. Bologna criteria are predictive for ovarian response and live birth in subsequent ovarian stimulation cycles. *Arch Gynecol Obstet.* 2019;299(2):571–7.

12. Gianaroli L, Racowsky C, Geraedts J, Cedars M, Makrigiannakis A, Lobo R. Best practices of ASRM and ESHRE: a journey through reproductive medicine. *Human Reprod (Oxford, England)*. 2012;27(12):3365–79.
13. Glujovsky D, Pesce R, Miguens M, Sueldo CE, Lattes K, Ciapponi A. How effective are the non-conventional ovarian stimulation protocols in ART? A systematic review and meta-analysis. *J Assist Reprod Genet*. 2020;37(12):2913–28.
14. Fischer S, Ehrig R, Schäfer S, Tronci E, Mancini T, Egli M, Ille F, Krüger THC, Leeners B, Röblitz S. Mathematical modeling and simulation provides evidence for new strategies of ovarian stimulation. *Front Endocrinol*. 2021;12:613048.
15. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol*. 2003;56(5):441–7.
16. Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: a review. *BMC Med*. 2010;8:20.
17. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, Altman DG, Moons KG. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med*. 2012;9(5):1–12.
18. McLernon DJ, Steyerberg EW, Te Velde ER, Lee AJ, Bhattacharya S. Predicting the chances of a live birth after one or more complete cycles of in vitro fertilisation: population based study of linked cycle data from 113 873 women. *BMJ (Clinical research ed)*. 2016;355:i5735.
19. Zegers-Hochschild F, Adamson GD, Dyer S, Racowsky C, de Mouzon J, Sokol R, Rienzi L, Sunde A, Schmidt L, Cooke ID, et al. The International Glossary on Infertility and Fertility Care, 2017. *Human Reprod (Oxford, England)*. 2017;32(9):1786–801.
20. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925–31.
21. Goodman LA. The analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odds. *Biometrics*. 1983;39(1):149–60.
22. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54(8):774–81.
23. Ratna MB, Bhattacharya S, Abdulrahim B, McLernon DJ. A systematic review of the quality of clinical prediction models in in vitro fertilisation. *Human Reprod (Oxford, England)*. 2020;35(1):100–16.
24. WHO WHO. WHO laboratory manual for the examination and processing of human semen. 5th ed 2010.
25. Morini D, Spaggiari G, Daolio J, Melli B, Nicoli A, De Feo G, Valli B, Viola D, Garganigo S, Magnani E, Pilia A, Polese A, Colla R, Simoni M, Aguzzoli L, Villani MT, Santi D. Improvement of sperm morphology after surgical varicocele repair. *Andrology*. 2021;9(4):1176–84. <https://doi.org/10.1111/andr.13012>.
26. Stolwijk AM, Wetzels AM, Braat DD. 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. *Human Reprod (Oxford, England)*. 2000;15(1):203–9.
27. Lintsen AM, Eijkemans MJ, Hunault CC, Bouwmans CA, Hakkaart L, Habbema JD, Braat DD. Predicting ongoing pregnancy chances after IVF and ICSI: a national prospective study. *Human Reprod (Oxford, England)*. 2007;22(9):2455–62.
28. Nelson SM, Lawlor DA. Predicting live birth, preterm delivery, and low birth weight in infants born from in vitro fertilisation: a prospective study of 144,018 treatment cycles. *PLoS Med*. 2011;8(1):e1000386.
29. Luke B, Brown MB, Wantman E, Stern JE, Baker VL, Widra E, Coddington CC 3rd, Gibbons WE, Ball GD. A prediction model for live birth and multiple births within the first three cycles of assisted reproductive technology. *Fertil Steril*. 2014;102(3):744–52.
30. Smith AD, Tilling K, Lawlor DA, Nelson SM. External validation and calibration of IVFpredict: a national prospective cohort study of 130,960 in vitro fertilisation cycles. *PLoS one*. 2015;10(4):e0121357.
31. Hemingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. *BMJ (Clinical research ed)*. 2009;339:b4184.
32. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361–87.
33. Santi D, Casarini L, Alviggi C, Simoni M. Efficacy of follicle-stimulating hormone (FSH) alone, FSH + luteinizing hormone, human menopausal gonadotropin or FSH + human chorionic gonadotropin on assisted reproductive technology outcomes in the “Personalized” medicine era: a meta-analysis. *Front Endocrinol*. 2017;8:114.
34. Paterlini M, Andrei F, Neri E, Trombini E, Santi S, Villani MT, Aguzzoli L, Agostini F. Maternal and paternal representations in assisted reproductive technology and spontaneous conceiving parents: a longitudinal study. *Front Psychol*. 2021;12:635630.
35. Scaravelli G, Zacà C, Levi Setti PE, Livi C, Ubaldi FM, Villani MT, Greco E, Coccia ME, Revelli A, Ricci G, Fusi F, Vigiliano V, De Luca R, Bolli S, Borini A. Fertilization rate as a novel indicator for cumulative live birth rate: a multicenter retrospective cohort study of 9,394 complete in vitro fertilization cycles. *Fertil Steril*. 2021;116(3):766–73. <https://doi.org/10.1016/j.fertnstert.2021.04.006>.
36. Shang J, Wang M, Wen Q, Ma Y, Chen F, Xu Y, Liu CH, Bai L, Tang H. A novel prognostic model to predict outcome of artificial liver support system treatment. *Sci Rep*. 2021;11(1):7510.
37. Imai K, Takai K, Miwa T, Maeda T, Hanai T, Shiraki M, Suet-sugu A, Shimizu M. Increased visceral adipose tissue and hyperinsulinemia raise the risk for recurrence of non-B non-C hepatocellular carcinoma after curative treatment. *Cancers (Basel)*. 2021;13(7):1542.
38. Park EY, Yi M, Kim HS, Kim H. A decision tree model for breast reconstruction of women with breast cancer: a mixed method approach. *Int J Environ Res Public Health*. 2021;18(7):3579.
39. Radhachandran A, Garikipati A, Zelin NS, Pellegrini E, Ghandian S, Calvert J, Hoffman J, Mao Q, Das R. Prediction of short-term mortality in acute heart failure patients using minimal electronic health record data. *BioData Min*. 2021;14(1):23.
40. Palmieri-Smith RM, Curran MT, Garcia SA, Krishnan C. Factors that predict sagittal plane knee biomechanical symmetry after anterior cruciate ligament reconstruction: a decision tree analysis. *Sports Health*. 2021;25:19417381211004932. <https://doi.org/10.1177/19417381211004932>.
41. Alves MA, Castro GZ, Oliveira BAS, Ferreira LA, Ramírez JA, Silva R, Guimarães FG. Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs. *Comput Biol Med*. 2021;132:104335.
42. van Eekelen R, Wang R, Danhof NA, Mol F, Mochtar M, Mol BW, van Wely M. Costeffectiveness of ovarian stimulation

- agents for IUI in couples with unexplained subfertility. *Hum Reprod.* 2021;36(5):1288–95. <https://doi.org/10.1093/humrep/deab013>.
43. Lee M, Lofgren KT, Thomas A, Lanes A, Goldman R, Ginsburg ES, Hornstein MD. The costeffectiveness of preimplantation genetic testing for aneuploidy in the United States: an analysis of cost and birth outcomes from 158,665 in vitro fertilization cycles. *Am J Obstet Gynecol.* 2021;225(1):55.e1–55.e17. <https://doi.org/10.1016/j.ajog.2021.01.021>.
 44. Neumann K, Griesinger G. An economic analysis of aneuploidy screening of oocytes in assisted reproduction in Germany. *Geburtshilfe Frauenheilkd.* 2020;80(2):172–8.
 45. Sadruddin S, Barnett B, Ku L, Havemann D, Mucowski S, Herrington R, Burggren W. Maternal serum concentration of anti-Müllerian hormone is a better predictor than basal follicle stimulating hormone of successful blastocysts development during IVF treatment. *PLoS one.* 2020;15(10):e0239779.
 46. van Eekelen R, Eijkemans MJ, Mochtar M, Mol F, Mol BW, Groen H, van Wely M. Cost-effectiveness of medically assisted reproduction or expectant management for unexplained subfertility: when to start treatment? *Hum Reprod.* 2020;35(9):2037–46.
 47. Xu H, Feng G, Wang H, Han Y, Yang R, Song Y, Chen L, Shi L, Zhang MQ, Li R, et al. A novel mathematical model of true ovarian reserve assessment based on predicted probability of poor ovarian response: a retrospective cohort study. *J Assist Reprod Genet.* 2020;37(4):963–72.
 48. Jing M, Lin C, Zhu W, Tu X, Chen Q, Wang X, Zheng Y, Zhang R. Cost-effectiveness analysis of GnRH-agonist long-protocol and GnRH-antagonist protocol for in vitro fertilization. *Sci Rep.* 2020;10(1):8732.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.