Running title: Predicting cumulative live birth after IVF

Predicting personalized cumulative live birth following in vitro fertilisation

David J McLernon, PhD[1], Edwin-Amalraj Raja, PhD[2], James P Toner, PhD[3], Valerie L Baker, MD[4], Kevin Doody, MD[5], David B Seifer, MD[6], Amy E Sparks, PhD[7], Ethan Wantman, MBA[8], Paul Lin, MD[9], Siladitya Bhattacharya, MD[10], Bradley J Van Voorhis, MD[11]

[1]Institute of Applied Health Sciences, Polwarth Building, University of Aberdeen, Aberdeen, AB25 2ZD, UK, d.mclernon@abdn.ac.uk;
[2]Institute of Applied Health Sciences, Polwarth Building, University of Aberdeen, Aberdeen, AB25 2ZD, UK, amalraj.raja@abdn.ac.uk;
[3]Department of Gynecology and Obstetrics, Emory University Atlanta, GA 30322, james.patrick.toner@emory.edu;
[4]Division of Reproductive Endocrinology and Infertility, Johns Hopkins University School of Medicine, Lutherville, Maryland 21093, valbaker@jhmi.edu;
[5]Center for Assisted Reproduction, 1701 Park Place Avenue, Bedford, TX 76022, kevind@embryo.net;
[6] Division of Reproductive Endocrinology & Infertility, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, david.seifer@yale.edu;
[7]Division of Reproductive Endocrinology and Infertility, University of Iowa Health Care, 1360 North Dodge Street, Iowa City 52245, amy-sparks@uiowa.edu;
[8]Redshift Technologies, Inc., 16 East 34th Street, New York, NY 10016, ethan@redshift-tech.com;
[9]Seattle Reproductive Medicine, 1505 Westlake Avenue North, Suite 400, Seattle, WA 98125, Paul.Lin@seattlefertility.com;
[10]Medical Sciences and Nutrition, Polwarth Building, University of Aberdeen, Aberdeen, AB25 2ZD, UK, s.bhattacharya@abdn.ac.uk;
[11]Division of Reproductive Endocrinology & Infertility, University of Iowa Health Care, 200 Hawkins Drive, 31325 PFP, Iowa City Iowa 52242, brad-van-voorhis@uiowa.edu.

Corresponding author: David McLernon

Address: Senior Research Fellow in Medical Statistics, Institute of Applied Health Sciences, Polwarth Building, University of Aberdeen, Aberdeen, AB25 2ZD, UK;
Email: d.mclernon@abdn.ac.uk
Tel: +441358 751248

Capsule
We developed two prediction calculators that provide the personalized probability of cumulative live birth before starting in vitro fertilisation and, in those whose first attempt failed, before the second stimulation.

ABSTRACT

Objective: To develop IVF prediction models to estimate the individualized chance of cumulative live birth at two time points: 1. Pre-treatment i.e. before starting the first complete cycle of IVF; and 2. Post-treatment i.e. before starting the second complete cycle of IVF in those couples whose first complete cycle was unsuccessful.

Design: Population-based cohort study.

Setting: National data from the Society for Assisted Reproductive Technology (SART) Clinic Outcome Reporting System.

Patients: 88,614 women who commenced IVF treatment using their own eggs and partner's sperm in SART member clinics.

Interventions: Not applicable.

Main Outcome Measures: The pre-treatment model estimated the cumulative chance of a live birth over a maximum of three complete cycles of IVF while the post-treatment model did so over the second and third complete cycles. One complete cycle included all fresh and frozen embryo transfers resulting from one episode of ovarian stimulation. We considered the first live birth episode including singletons and multiples.

Results: Pre-treatment predictors included female age (35 vs 25 years, adjusted odds ratio 0.69, 95% confidence interval 0.66 to 0.73) and BMI (35 vs 25 kg/m^2, 0.75, 0.72 to 0.78). The post-treatment model additionally included number of eggs from the first complete cycle (15 vs 9 eggs, 1.10, 1.03 to 1.18).

Using the pre-treatment model, a nulliparous woman aged 34 years, BMI=23.3, male partner infertility and AMH=3ng/ml has a 61.7% chance of having a baby over her first complete cycle of IVF (cumulatively, over three complete cycles=88.8%). If this is unsuccessful, using the post-treatment model her chance of live birth over the second complete cycle 1 year later (age=35 years, number of eggs =7) is 42.9%. The C-statistic for all models was between 0.71 and 0.73.

Conclusions: The focus of previous IVF prediction models based on American data have been live birth after treatments involving either fresh or frozen embryos. These novel prediction models provide clinically relevant estimates which could help clinicians and couples plan IVF treatment at different points in time.

Keywords: IVF, prediction, cumulative live birth

INTRODUCTION

To date, over eight million babies have been born worldwide as a result of in-vitro fertilisation (IVF) (1). The number of treatments continue to rise, and IVF now accounts for approximately 1.8% of all live born babies in the USA (2,3).

Until recently, IVF success rates have been reported in terms of live birth rate following a single cycle ending in the transfer of one or more fresh embryos. Improvements in embryo cryopreservation techniques and the subsequent increase in frozen-thawed embryo transfers (FET) have encouraged clinicians to report cumulative live birth rates resulting from the transfer of fresh followed by frozen embryos as a more comprehensive measure of success (4-7). A complete cycle of IVF now includes the initial "fresh" embryo transfer as well as the additional contributions from the transfer of frozen embryos (FETs) derived from oocytes retrieved from the same episode of ovarian stimulation. In addition, as many women undergo several IVF treatments, each involving ovarian stimulation followed by the replacement of embryos generated from the oocytes retrieved, estimating outcomes from multiple complete cycles is critical in terms of providing patients with a personalized chance of success over a series of attempts (8).

Previously developed clinical prediction models have estimated the individualized predicted probability of cumulative live birth over three fresh embryo transfer attempts (9), but have excluded the associated FETs which allow us to report chances of success over one or more complete cycles (10). A UK model is able to predict cumulative live birth over multiple complete cycles of IVF but cannot predict outcomes after a failed first complete cycle (11,12). Crucially, the model lacks potentially key predictors such as female body mass index (BMI) and markers of ovarian reserve such as anti-Mullerian hormone (AMH) and follicle stimulating hormone (13).

In this study, we aimed to develop two more advanced versions of these models. The first was a "pre-treatment" model that could predict the cumulative probability of live birth over three complete cycles of IVF and the second was a "post-treatment" model to revise predictions of success in couples whose first complete cycle was unsuccessful.

MATERIALS AND METHODS

This project was deemed not to require review by the University of Iowa Institutional Review Board because the study is limited to de-identified data from an external database.

Database

The Society for Assisted Reproductive Technology (SART) collects fertility treatment data from over 90% of reported IVF treatment cycles in the USA. Data on all women who commenced IVF treatment in SART member clinics were stored in the SART Clinical Outcome Reporting System (SART CORS) database (http://www.sart.org/research/). The SART CORS data were collected and verified by SART and reported to the Centers for Disease Control and Prevention in compliance with the Fertility Clinic Success Rate and Certification Act of 1992 (Public Law 102-493). The data are validated annually with some clinics having on-site visits for chart review based on an algorithm for clinic selection. During each visit, data reported by the clinic were compared with information recorded in patient charts. Ten out of 11 data fields selected for validation were found to have discrepancy rates of ≤5% (14).

Beginning with cycles starting on or after January 1st, 2014, the SART CORS collected retrieval information for each initiated FET cycle, allowing previously entered attempts at retrieval to be "linked" to each FET. Included in the dataset which was securely sent to DJM, AR and SB, were identifiers that allowed linkage of FET cycles to retrieval cycles. This included diagnostic and treatment characteristics of those source cycles (attempts at retrieval) that were found to be linked with one or more FET cycles for a given woman.

SART CORS provides a research portal for members to request de-identified datasets. These requests are managed by the SART Research Committee (volunteer group of physicians and embryologists who are elected and appointed) in a portal that tracks requests, approvals / rejections, communications, dissemination of data, and resulting publications.

De-identification of the SART CORS data was performed in accordance with the Health Insurance Portability and Accountability Act of 1996 (HIPAA). The de-identified dataset was transferred to DJM, AR and SB at the University of Aberdeen in an encrypted file.

Study population

This population-based cohort study included all women who commenced IVF treatment using their own eggs and partner's sperm in SART member clinics. Data were extracted for all women who started their first autologous IVF cycle during 2014-2015. The linked nature of the dataset allowed us to restructure it into complete cycles per woman within a clinic, where a complete cycle is defined as all fresh and frozen embryo transfer attempts resulting from one episode of ovarian stimulation. This definition also includes women who did not have any eggs retrieved, either due to a cancelled cycle or not having any eggs available at aspiration. We were unable to identify women who had treatment at

more than one clinic. However, we expect this number to be relatively low over a two-year time period. We included all associated FET attempts that were carried out during 2016 in order to capture at least one year of follow-up within a complete cycle. Women who had IVF treatment over a maximum of three complete cycles were included. The study followed women until the complete cycle which resulted in their first live birth or until the last complete cycle (up to the maximum of three) that did not result in a live birth.

The main reasons women were excluded from the study were:

(i) they were not between 18 and 50 years of age at the start of their first IVF treatment;
(ii) they banked embryos or oocytes for the purposes of fertility preservation;
(iii) they had an embryo transfer using banked embryos and banked oocytes which were for the purpose of fertility preservation;
(iv) genetic testing of embryos was performed;
(v) they had more than 3 embryos transferred in any of the treatment cycles;
(vi) they had embryos transferred to the uterus which originated from different egg retrievals;
(vii) the treatment involved egg donation, sperm donation, or a gestational carrier.

A list of the exclusions and numbers of women excluded are in Figure S1.

Baseline characteristics

The baseline characteristics of couples included female age, previous full-term birth (no versus yes), male factor infertility (yes versus no), endometriosis (yes versus no), tubal factor (yes versus no), uterine factor (e.g. septum, myoma, intrauterine adhesions, congenital anomalies) (yes versus no), polycystic ovaries or polycystic ovary syndrome (PCOS) (yes versus no), diminished ovarian reserve (yes versus no), unexplained infertility (yes versus no), body mass index, and serum anti-Mullerian hormone concentration (AMH). We also recorded the number of eggs retrieved during the first complete cycle.

Outcome

The cumulative predicted probability of a live birth for a couple up to three complete cycles of IVF over a maximum of three years. One complete cycle included all fresh and frozen embryo transfers resulting from one episode of ovarian stimulation. The live birth episode included singletons and multiples and we only considered the first IVF live birth.

Statistical methods

Descriptive statistics were presented for each predictor. Continuous data were presented using the mean and standard deviation (SD) and categorical data were presented as frequency and percentage.

Clinical prediction models were developed for two different time points:

Pre-treatment model: For a couple about to embark on IVF, we estimated the cumulative probability of live birth over a maximum of three complete cycles. We used the characteristics of the couple just before egg retrieval as predictors. We developed two versions of this model. One for women who did not have an AMH measurement (AMH included as a predictor), and one for women who did.

Post-treatment model: For couples who wish to have further IVF treatment after a first complete cycle, we developed a model that could revise the cumulative predicted probability of live birth over the next two complete cycles. We used updated couple characteristics before the second complete cycle as predictors as well as the number of eggs collected in the first cycle (including zero eggs). We did not include AMH in this model because too few women in the dataset had AMH measurements at the second complete cycle to reliably predict live birth.

Development of pre-treatment model for predicting live birth from the first complete cycle

A discrete time logistic regression model was used to predict the chance of a live birth over a maximum of three cumulative complete cycles of IVF. A complete cycle included a fresh embryo transfer and any associated frozen-thawed embryo transfers, or, for women who chose to freeze all embryos, all associated frozen-thawed embryo transfers. If a woman only had one or two complete cycles and these were unsuccessful she was censored, meaning that either the study period ended before she could start a new cycle, she was lost to follow-up or she genuinely stopped treatment. We were unable to distinguish between these different reasons for censoring. The predicted probability of a live birth in the *i*th complete cycle is estimated conditional on no live birth having occurred before that complete cycle. From this model, we calculated the cumulative probability of a live birth over sequential complete cycles up to complete cycle 3. A separate model was fitted for those women who had an AMH measurement.

The woman's age, BMI and AMH at the start of first cycle had non-linear relationships with the outcome and these were accounted for using restricted cubic spline functions in the model. Since spline terms are difficult to interpret clinically, we estimated the effect for different values of each predictor relative to a reference value. To avoid the effect of influential extreme values we truncated BMI and AMH to the 99th percentile value before fitting the model.

A manual backward selection process (eliminating predictors until the Akaike's Information Criterion was at its lowest) was used to determine the final model to

predict live birth over successive complete cycles of IVF. In this procedure, complete cycle number was always included, as was the woman's age since it is a known predictor of pregnancy outcomes after IVF (13). All of the other available predictors were subjected to the selection process. When we added AMH to the model, we kept all the predictors from the model without AMH.

Development of post-treatment model for predicting live birth from the second complete cycle

In women whose first complete cycle did not result in a live birth, a similar model was developed to predict the cumulative chance of a live birth over the subsequent two complete cycles of IVF. All values of predictors, such as age, BMI and diagnosis, were revised to reflect the current status of the patients just before the start of the second complete cycle. One exception was the number of eggs collected as these were recorded from the first complete cycle. Woman's age, BMI and the number of eggs retrieved at the first complete cycle were transformed using a restricted cubic spline function. A similar manual backward selection process to the one used to develop the pre-treatment was conducted to determine the final post-treatment model.

Missing data

Where missing data were present for any predictor, the characteristics of women with complete data were compared to those with missing data. Single imputation was performed to impute values for BMI and full-term live birth status. This procedure assumes that missing data were missing at random, conditional upon the observed predictors and outcome (see Text S1 for further details).

Predictive ability

To detect possible overfitting of the model we calculated a heuristic shrinkage factor, *s*. The shrinkage factor was calculated from a heuristic formula: $s = (\text{model chi}^2 - \text{df})/\text{model chi}^2$,

where model chi^2 is the likelihood ratio of the model and df is the degrees of freedom of the number of candidate predictors considered for the model. If *s* is very close to 1 then there is no evidence of overfitting (15).

To assess the predictive ability of the models, discrimination and calibration statistics were estimated. The ability of the models to discriminate between couples at high and low chances of a live-birth were assessed using Uno's C-statistic (95% CI) which ranges between 0.5 and 1 (16). A C-statistic of 1 indicates perfect discrimination, whereas a C-statistic of 0.5 represents a model with no discrimination at all. Calibration plots were created showing agreement between the observed live birth and predicted cumulative probability of live birth for each complete cycle. These were developed by splitting the predicted probability of live birth into tenths and calculating the average probability for

each cumulative cycle within each tenth. The observed proportion with live birth for each cumulative cycle was calculated within each tenth using the Kaplan-Meier method.

All analyses were performed using Stata MP version 15 and SAS version 9.4 (SAS Institute).

RESULTS

After exclusions, 88,613 couples who underwent 121,561 complete cycles of IVF were included in the analysis (see Fig S1). Table 1 shows the distribution of the characteristics of the couples at the start of their first complete cycle which resulted in a live birth in 40,707 (45.9%) women (Fig S2). Of the remaining 47,906 couples, 24,735 (51.6%) went on to have a second complete cycle and 7,214 (29.2%) had a live birth. Of the remaining 17,523 couples, 8,213 (46.9%) went on to have a third complete cycle resulting in a live birth in 1,618 (19.7%). Overall, 55.9% of all couples had a live birth over the first three complete cycles of IVF during the three-year study period.

BMI was missing for 18,246 (20.6%) women (see Table S1) and previous full-term live birth was missing for 333 (0.4%) women. These data were imputed as detailed in supplementary text S1.

Pre-treatment model: Predicting cumulative live birth from first complete cycle

Figure S3 shows the unadjusted relationship between the predicted probability of live birth in the first complete cycle and the woman's age and BMI. The chances of a live birth declined after age 30. Across the range of BMI, the range of average predictions is not very wide (~35-50%), peaking at a BMI of around 22. Rising serum AMH levels displayed a steep increase in the predicted probability of live birth up to 2ng/mL, beyond which the relationship began to weaken.

The factors that predicted cumulative live birth were female age and BMI, previous full-term birth, male factor infertility, polycystic ovaries/PCOS, diminished ovarian reserve, uterine factor infertility and unexplained infertility (see Table 2). Predictors which were not retained in the final model included tubal factor and endometriosis. The odds of live birth decreased with every additional complete cycle e.g. the odds of live birth after complete cycle 2 were 42% lower than the odds after complete cycle 1. The odds of live birth decreased with increasing female age e.g. for women aged 35 versus women aged 25, Odds Ratio (OR) = 0.69 (95% CI 0.66 to 0.73). Table 2 provides odds ratios for different ages based on the restricted cubic spline curve (see Fig S3A for the non-linear relationship between age and predicted probability of live

birth). Women who had a previous full-term birth had slightly higher odds of live birth than women who did not (OR = 1.05 (1.01 to 1.08). Couples with male factor infertility, polycystic ovaries/PCOS or unexplained infertility had higher odds of live birth whereas couples with uterine factor infertility or diminished ovarian reserve had lower odds of live birth. Higher female BMI was associated with decreased chance of live birth. Women with a BMI of 30 had 14% lower odds of live birth compared to women with a BMI of 25 (OR = 0.86 (95% CI 0.84 to 0.89).

For the model including serum AMH levels, the effects of the other predictors were similar to those in the model without this information. However, the effect of polycystic ovaries/PCOS was weaker. AMH was a significant predictor and as the values of AMH increased so did the odds of live birth e.g. a woman with an AMH of 5 had 22% increased odds of live birth compared to a woman with an AMH of 2.5.

Post-treatment model: Predicting cumulative live birth from second complete cycle

Table 1 shows the distribution of the characteristics of those couples who started a second complete cycle after an unsuccessful first complete cycle. Compared to women starting their first complete cycle, there was a higher distribution of diminished ovarian reserve (38.9% versus 24.5%) and PCOS/polycystic ovaries (38.2% versus 27.4%) in this subset of women starting their second complete cycle.

Higher numbers of retrieved eggs from the first complete cycle was associated with increased odds of live birth e.g. a woman with 15 oocytes retrieved had 10% increased odds of live birth compared to a woman who had 9 oocytes retrieved, OR = 1.10 (1.03 to 1.18) (see Table 2 for odds ratios for different numbers of eggs and Fig S3D for the graphical relationship between egg number and predicted live birth). The effects of the other predictors in the post-treatment model were similar to those in the pre-treatment model.

Assessing ability to predict

The performance statistics are presented in Table S2. The heuristic shrinkage factor was almost 1 for all models signifying no overfitting. The C-statistic for the pre-treatment model was 0.71 and this increased to 0.73 when AMH was included in the model. For the post-treatment model, the C-statistic as 0.71. Figure 1 shows the calibration plots for different numbers of cumulative complete cycles for the pre- and post-treatment models. All tenths of predicted probability of live birth are in good agreement with the corresponding proportion of observed live births.

Examples of the models in hypothetical couples

The model equations are contained in Supplementary Text S2. To demonstrate the models, we plotted the predicted probability of live birth over the first complete cycle for hypothetical women with a BMI of 24, no previous live birth and with unexplained infertility who are about to start their first IVF cycle. We plotted their predicted probabilities for different ages and for the range of AMH from 0 to 16 (Figure 2). For example, a woman aged 25 and AMH of 5 has a 67.5% chance of a live birth over her first complete cycle of IVF (using pre-treatment model). However, a woman aged 40 and AMH of 1 has a 26.2 chance of live birth. Cumulatively, over three complete cycles their chances of having a live birth are 39.2% and 47.7% respectively. Assuming that the first complete cycle is unsuccessful and these women continue IVF 2 years later (when their ages have increased to 27 and 42 and the number of eggs retrieved in the first cycle was 10 and 3 respectively) their chance of having a live birth over the second complete cycle (using post-treatment model) becomes 48.4% and 12.4% respectively. Cumulatively, over the second and third complete cycles their chances are 68.8% and 20.3% respectively.

We have developed an online calculator that clinicians and couples can use to calculate their own predicted probability of live birth (available at sart.org or https://w3.abdn.ac.uk/clsm/SARTIVF/).

DISCUSSION

We have developed two novel clinical prediction models that estimate the personalized cumulative probability of having a baby over the first three complete cycles of IVF. The first can be used by treatment naïve couples while the second uses additional information from a first IVF cycle to predict chances of future success for couples considering further therapy. Both models were shown to predict accurately in the USA population.

Strengths and weaknesses

The models were developed using a large national dataset which included over 90% of the reported IVF cycles performed in the USA providing increased accuracy for prediction and generalizability. They are the first models in the USA to predict the probably of live birth for couples with different causes of infertility over complete IVF cycles which include the contribution of frozen-thawed embryo transfers. The pre-treatment model is one of the first to include serum AMH, the most sensitive and earliest biomarker of a decline in ovarian reserve, as a predictor (17). The post-treatment model is the first of its kind to provide revised predictions for couples who have completed an initial IVF cycle which was unsuccessful.

Some potentially important predictors including duration of infertility, ethnicity, serum basal follicle stimulating hormone (FSH) and smoking status were missing from sufficient numbers of records to make them unusable for our model. However, we used AMH instead of FSH as it is seen as a more reliable predictor of ovarian reserve (18-21). While smoking status and alcohol consumption have an association with IVF success (22-24), the self-reported nature of these variables questions their quality (25,26). The heterogeneity of data on race and ethnicity (missing in around 40% of records) in this dataset have been described elsewhere (5,9).

Unfortunately, we were unable to assess the impact of the variable accuracy of different AMH assays used by SART CORS reporting clinics in our analysis. During the study period of 2014-2015 the most likely AMH assay employed was the manual Gen2 Beckman-Coulter ELISA assay. Other clinics would have been using laboratory derived assays such as Quest or LabCorp AMH test. However, information such as the specific AMH assay or laboratory where AMH was performed (i.e. in house or commercial laboratory) was not recorded in the SART CORS database. Despite the 'noise' introduced by different assays, AMH is a strong predictor of cumulative live birth which confirms the utility of this measure in our models using SART CORS data (17). The currently available automated testing platforms (based on electrochemiluminescence technology) have a much greater consistency and will likely be even more strongly correlated as we update the predictor model with new data in the future.

We were unable to identify women who had treatment at more than one clinic. The SART CORS database does not have a way of linking women identifiers across different clinics.

Our models were limited to predicting live birth for heterosexual couples undergoing autologous IVF as the data from the years 2014-16 did not have sufficient numbers of single women or same-sex couples to allow development of an accurate prediction model for these groups. The relatively short follow-up time period also meant that we could only develop a post-treatment model for couples whose first complete cycle attempt was unsuccessful. Since all complete cycles had to begin during 2014 and 2015, few couples would have embarked on a second complete cycle after having a baby in that time. 51.6% of women who did not have a live birth in the first complete cycle went on to commence a second. A similar proportion who failed the second complete cycle started a third. This lower than expected rate is attributable to a combination of genuine discontinuation of treatment and administrative censoring caused by the relatively short study period. Nevertheless, our calibration assessments showed that there were enough live births in the third complete cycle to provide accurate cumulative predictions over three cycles. However, the precision was slightly poorer meaning that there may be more uncertainty for cumulative predictions over three complete cycles.

Previous studies

The last three decades have seen the publication of many IVF prediction models, most of which are only able to estimate live birth outcomes following individual fresh embryo transfer episodes with no consideration of FETs (10,27,28). However, four published prediction models were found that are worthy of further discussion as they consider the contribution of FETs and/or predict over multiple cycles of IVF (9,11,17,29). The van Loendersloot model calculates pregnancy chances for individual fresh and frozen embryo transfers (29). If an embryo transfer does not result in a pregnancy, the model can revise the predicted probability for the following embryo transfer. However, it cannot be used to estimate cumulative chances over complete IVF cycles, does not have BMI and serum AMH as predictors, and was developed and validated using data from one Dutch clinic. The Luke models used US data but only makes cumulative predictions using fresh embryo transfer attempts whilst excluding any associated frozen embryo transfers which is not usual clinical practice in IVF (9). Given the integral role of cryopreservation techniques in IVF and the associated increased in frozen-thawed embryo transfer, our complete cycle definition provides a more clinically and contemporary relevant estimate of the chance of live birth over an entire course of IVF treatment. Further, the Luke study developed separate models according to the number of embryos transferred meaning that these models can only be used to make predictions after embryo transfer - at the end of a fresh IVF cycle rather than at the start. Unlike our study which analysed these predictors as continuous variables, Luke et al categorised age and BMI in their models which results in reduced precision and power and ignores the relationship that exists between predictor values and the outcome within categories (30,31). Furthermore, the Luke models did not include serum AMH values. The Tal model predicts cumulative live birth over complete cycles of IVF but only in women who have diminished ovarian reserve and only from the first complete cycle (17).

The UK-based McLernon models were developed using the Human Fertilisation and Embryology Authority (HFEA) national registry which contains data on fresh and frozen IVF cycles linked to individual women from the UK (11). Like the pre-treatment model in this study, the UK version predicts cumulative live birth over multiple complete cycles of IVF before treatment starts. However, unlike our model, the UK model did not contain important predictors such as BMI or AMH since these were not available in the HFEA registry. Furthermore, the UK model does not revise cumulative predictions after an unsuccessful first complete cycle.

Our new models showed good predictive performance upon internal validation without any evidence of overfitting. The discrimination ability for the pre-treatment model is similar to that found in the UK model which had a C-statistic of 0.69, slightly lower than our C-statistic of 0.71 (or 0.73 with the addition of

serum AMH). The Luke models were not assessed for discrimination ability. To interpret the C-statistic, we need to imagine all possible pairs of couples from the dataset within which one couple has a live birth and the other does not. A C-statistic of 0.71 means that in 71% of these pairs, the model correctly assigns a higher predicted probability of live birth to the couple who had a live birth. In reproductive medicine prediction models, discrimination performance frequently results in C-statistics of between ~0.6 to ~0.7. It has been suggested that this is due to the homogeneity of the population being studied (32).

Interpretation and clinical importance

Our two models (and their AMH variants) may be used by clinicians and couples at two different time points. The pre-treatment model is applied before the first IVF cycle begins and can provide cumulative predictions over the first three complete cycles. Since couples often must pay out of pocket for ART treatment in the USA, they may use the pre-treatment models to help them make a decision on whether or not it is cost-effective for them to embark on IVF. For those whose first complete cycle did not end in a live birth, the post-treatment model is applied before the start of the second complete cycle and provides cumulative predictions over the second and third complete cycles. This model may be useful to help couples make a decision on whether to continue with IVF using the extra treatment information learned from the first attempt. Such information should help to empower couples to plan their treatment and prepare emotionally and financially at different stages of their IVF treatment. Predictions are likely to be more precise for the first complete cycle since there is likely to be more complete capture of all treatments that occurred over the first complete cycle. Although the model performed well, the cumulative prediction over three complete cycles is likely to have slightly weaker precision. This is because there was a high proportion of censoring in the relatively short follow-up period that was available in the SART CORS dataset.

The models have been converted into an easy to use online calculator which is available for use by clinicians and couples at sart.org. Let's describe an example of an application of the pre-treatment model by considering a 34-year-old woman who is 140 pounds and 5 feet 5 inches tall. She has no cause of infertility, but her partner has a problem with his sperm. She has never had a baby born to full-term and her AMH level is 3ng/ml. The couple have decided to have IVF treatment and when they enter the above information into the calculator, they find that their chance of having their first baby after one complete cycle is 61.7. Their cumulative chance over two complete cycles is 80.6%. Unfortunately, their first complete cycle was unsuccessful, but they have decided to try IVF again. They enter their updated information into the post-treatment calculator: The woman is now 35 years old and she had 7 eggs collected in her first ever IVF cycle. The couple find that their chance of having

their first baby after this second complete cycle is 42.9%, and cumulatively over the second and third is 62.4%.

Further research

Our models were developed using all complete cycles of IVF started over a two-year period plus one year for associated FETs. Future work will involve temporal validation and, if necessary, subsequent updating of the models using new extracts of SART CORS data over a longer period. This will allow us to calculate more precise and reliable estimates of national level cumulative live birth rates especially for the third complete cycle.

Additional data will also allow us to make predictions for subgroups of patients such as single women and same-sex couples as well as a post-treatment model for women who wish to try for a second IVF baby.

The SART CORS dataset includes information on over 90% of reported IVF cycles performed in the USA. Such a large dataset allows accurate predictions at an average national level. However, we note that individual clinics may have their own historical patient level datasets which may be used to develop clinic specific prediction models. This may result in predictions that are more specific to their patient population than our SART CORS models. Rather than develop a brand-new model, we recommend that clinics first validate our models using their datasets. If performance is shown to be poor, then the models can be updated using recommended statistical methods (33).

Potentially important predictors that we found to be poorly recorded, such as race and duration of infertility, may be better recorded in future data extracts. This would enable their inclusion in the above proposed models and in future iterations of existing models to further refine predictions.

It is important that future research also focuses on the utility of IVF prediction models and the perceptions of couples who have used them. This will allow us to further refine our models and their online presentation.

CONCLUSIONS

We have developed two novel prediction models that can provide clinically relevant individualised estimates of the cumulative chances of live birth before starting IVF and, subsequently, in those whose first attempt failed, before the second stimulation. These models have been converted into an online calculator which could help clinicians and couples plan IVF treatment at different points in time.

Acknowledgements

SART wishes to thank all its members for providing clinical information to the SART CORS database for use by patients and researchers. Without the efforts of our members, this research would not have been possible. We thank Barbara Luke and Morton Brown for cleaning and coding the dataset for analysis before sending to DJM, AR and SB.

Contributors

DJM, SB, BVV designed the study. AR and DJM conducted the statistical analysis. DJM did the literature search and wrote the article. All authors contributed intellectually to the writing or revising of the manuscript and approved the final version. DJM is the guarantor.

Funding

This project was funded by SART.

Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

REFERENCES

1. Adamson GD, Zegers-Hochschild F, Dyer S, Chambers G, Ishihara O, Mansour R, et al. International Committee for Monitoring Assisted Reproductive Technologies (ICMART) World Report on ART, 2014 and The International Glossary on Infertility and Fertility Care, 2017. Abstracts of the 34th Annual Meeting of the European Society of Human Reproduction and Embryology. Hum Reprod 2018;33(Suppl 1):i65.

2. Sunderam S, Kissin DM, Zhang Y, Folger SG, Boulet Sl, Warner L, et al. Assisted Reproductive Technology Surveillance — United States, 2016. MMWR Surveill Summ 2019;68(4).

3. Toner JP, Coddington CC, Doody K, Van Voorhis B, Seifer DB, Ball GD, et al. Society for Assisted Reproductive Technology and assisted reproductive technology in the United States: a 2016 update. Fertil Steril 2016;106:541-6.

4. Malizia BA, Hacker MR, Penzias AS. Cumulative live–birth rates after in vitro fertilization. N Engl J Med 2009;360:236–43.

5. Luke B, Brown MB, Wantman E, Lederman A, Gibbons W, Schattman GL, et al. Cumulative birth rates with linked assisted reproductive technology cycles. N Engl J Med 2012;366:2483–91.

6. Paul RC, Fitzgerald O, Lieberman D, Venetis C, Chambers GM. Cumulative live birth rates for women returning to ART treatment for a second ART-conceived child. Hum Reprod 2021;35:1432-40.

7. Maheshwari A, McLernon D, Bhattacharya S. Cumulative live birth rate: time for a consensus? Hum Reprod 2015;30:2703–7.

8. McLernon DJ, Maheshwari A, Lee AJ, Bhattacharya S. Cumulative live birth rates after one or more complete cycles of IVF: a population–based study of linked cycle data from 178,898 women. Hum Reprod 2016;31:572–81.

9. Luke B, Brown MB, Wantman E, Stern JE, Baker VL, Widra E, et al. A prediction model for live birth and multiple births within the first three cycles of assisted reproductive technology. Fertil Steril 2014;102:744-52.

10. Ratna M, Bhattacharya S, Abdulrahim B, McLernon DJ. A systematic review of the quality of clinical prediction models in in vitro fertilisation. Hum Reprod 2020;35:100-16.

11. McLernon DJ, te Velde E, Steyerberg E, Lee AJ, Bhattacharya S. Predicting the chances of a live birth after one or more complete cycles of in-vitro fertilisation: an population-based study of linked cycle data from 113,873 women. BMJ 2016;355:i5735.

12. Leijdekkers JA, Eijkemans MJC, van Tilborg TC, Oudshoorn SC, McLernon DJ, Bhattacharya S, et al. Predicting the cumulative chance of live birth over multiple complete cycles of in vitro fertilisation: an external validation study. Hum Reprod 2018;33:1684-95.

13. Van Loendersloot LL, van Wely M, Limpens J, Bossuyt PMM, Repping S, van der Veen F. Predictive factors in in vitro fertilisation (IVF): a systematic review and meta–analysis. Hum Reprod Update 2010;16:557–89.

14. Center for Disease Control and Prevention, American Society for Reproductive Medicine, and Society for Assisted Reproductive Technology. Assisted reproductive technology fertility clinic success rates report. Washington, DC: US Dept. of Health and Human Services; 2016. p. 523-5. Available at: https://www.cdc.gov/art/reports/archive.html. Accessed 17th May 2021.

15. Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-87.

16. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med 2011;30:1105–1117.

17. Tal R, Seifer DB, Tal R, Granger E, Wantman E, Tal O. AMH Highly Correlates with Cumulative Live Birth Rate in Women with Diminished Ovarian

Reserve Independent of Age. J Clin Endocrinol Metabol 2021; https://doi.org/10.1210/clinem/dgab168

18. Tal R, Seifer DB. Ovarian reserve testing: a user's guide. Am J Obstet Gynecol 2017;217:129-140.

19. Toner JP, Seifer DB. Why we may abandon basal follicle-stimulating hormone testing: a sea change in determining ovarian reserve using antimullerian hormone. Fertil Steril 2013;99:1825-30.

20. Barad DH, Weghofer A, Gleicher N. Comparing anti-Müllerian hormone (AMH) and follicle-stimulating hormone (FSH) as predictors of ovarian function. Fertil Steril 2009;91:1553-5.

21. Seifer DB, MacLaughlin DT, Christian BP, Feng B, Shelden RM. Early follicular serum mullerian-inhibiting substance levels are associated with ovarian response during assisted reproductive technology cycles. Fertil Steril 2002;77:468-71.

22. Firns S, Cruzat VF, Keane KN, Joesbury KA, Lee AH, Newsholme P, et al. The effect of cigarette smoking, alcohol consumption and fruit and vegetable consumption on IVF outcomes: a review and presentation of original data. Reprod Biol Endocrinol 2015;13:134.

23. Rossi BV, Berry KF, Hornstein MD, Cramer DW, Ehrlich S, Missmer SA. Effect of alcohol consumption on in vitro fertilization. Obstet Gynecol 2011;117:136-42.

24. Soares SR, Melo MA. Cigarette smoking and reproductive function. Curr Opin Obstet Gynecol 2008;20:281-91.

25. Boniface S, Kneale J, Shelton N. Drinking pattern is more strongly associated with under-reporting of alcohol consumption than socio-demographic factors: evidence from a mixed-methods study. BMC Public Health 2014;14:1297.

26. Håkonsen LB, Ernst A, Ramlau-Hansen CH. Maternal cigarette smoking during pregnancy and reproductive health in children: a review of epidemiological studies. Asian J Androl 2014;16:39-49.

27. 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:e1000386.

28. Dhillon RK, McLernon DJ, Smith PP, Fishel S, Dowell K, Deeks JJ, et al. Predicting the chance of live birth for women undergoing IVF: a novel pretreatment counselling tool. Hum Reprod 2016;31:84-92.

29. van Loendersloot LL, van Wely M, Repping S, Bossuyt PMM, van der Veen F. Individualized decision-making in IVF: calculating the chances of pregnancy. Hum Reprod 2013;28:2972-80.

30. Altman DG. Suboptimal analysis using 'optimal' cutpoints. Brit J Cancer 1998;78:556-7.

31. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using 'optimal' cutpoints in the evaluation of prognostic factors. J Nat Cancer Inst 1994;86:829-35.

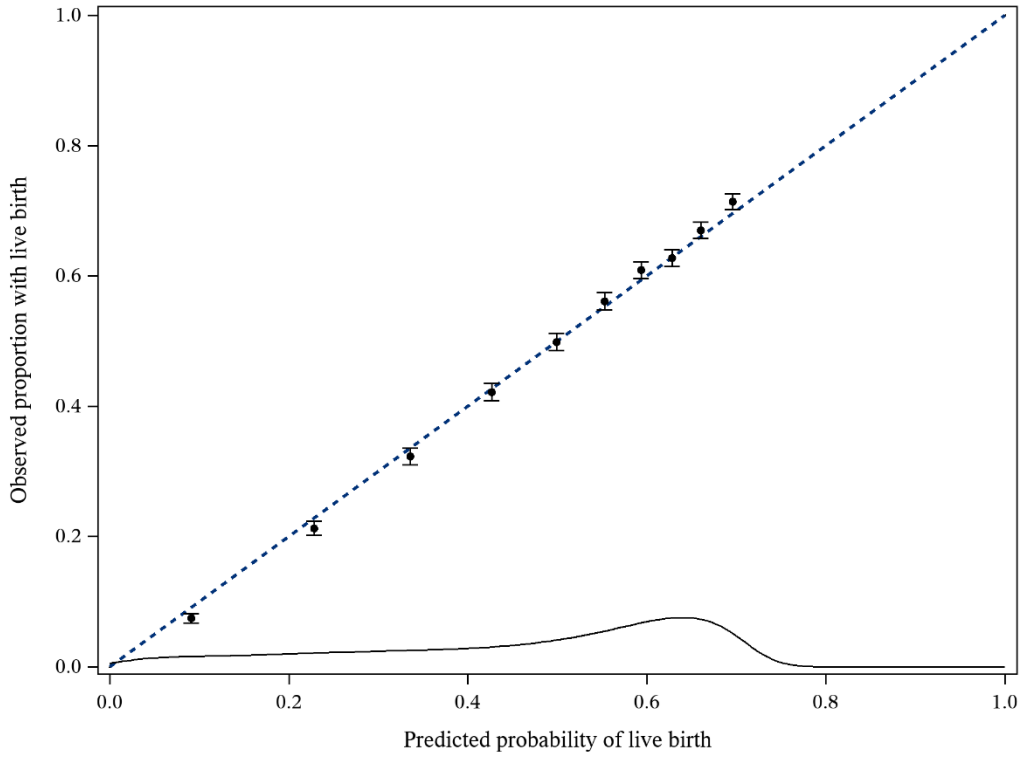
32. Coppus SFPJ, van der Veen F, Opmeer BC, Mol BWJ, Bossuyt PMM. Evaluating prediction models in reproductive medicine. Hum Reprod 2009;24:1774-8.

33. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol 2008;61:76-86. doi: 10.1016/j.jclinepi.2007.04.018.

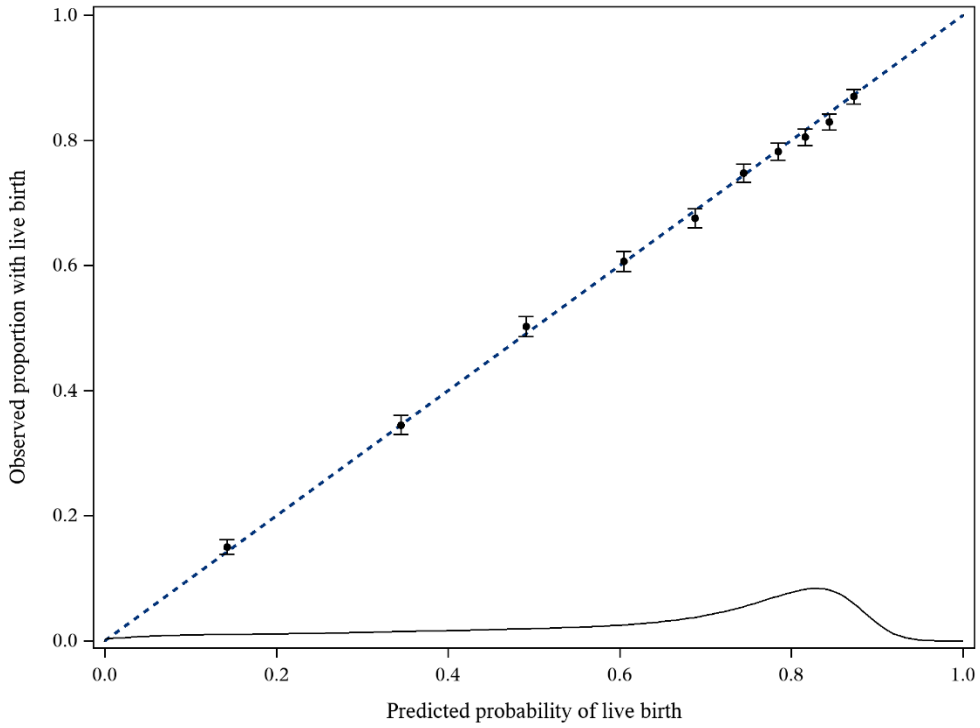
FIGURE CAPTIONS

Figure 1 Calibration plots showing the observed cumulative proportion of live birth (95% CI) from the development dataset and the predicted cumulative probability of live birth over: A the first complete cycle for pre-treatment model (including AMH), B two complete cycles for pre-treatment model (including AMH) C the second complete cycle for post-treatment model; and D over the second and third complete cycles for post-treatment model. The density curve shows the distribution of predictions. The ten points on the graphs were created as follows: the predicted probabilities for all patients were ranked in order and then grouped into tenths. Within each tenth the averaged predicted probability of live birth from the model was plotted against the observed proportion of live births (estimated from Kaplan-Meier with 95% confidence intervals plotted). The dashed diagonal line reflects perfect agreement.

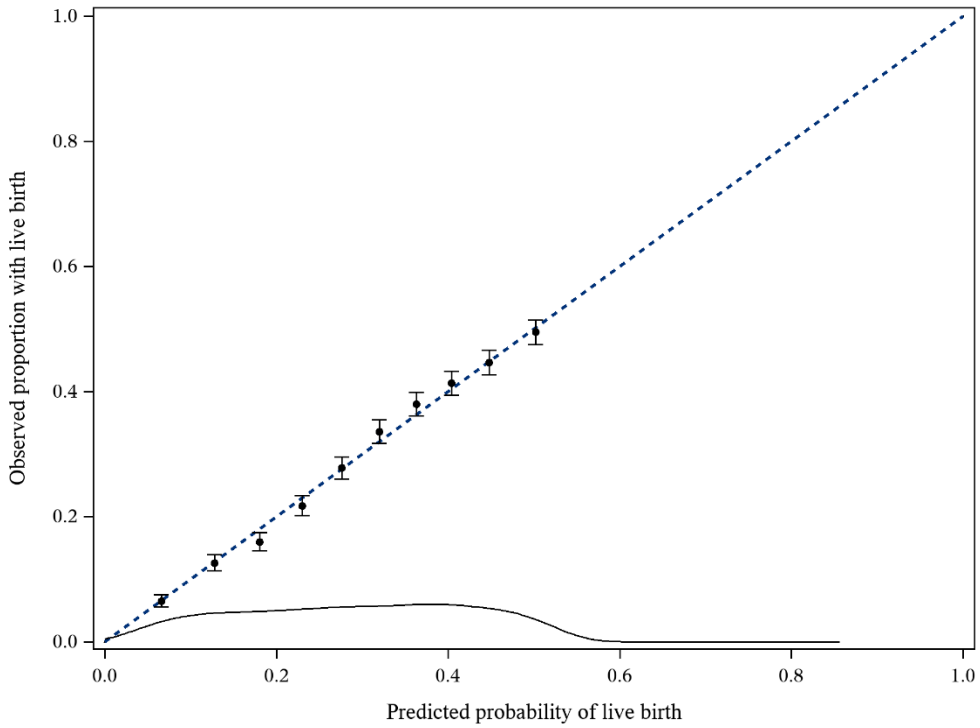
A



B


Observed proportion with live birth
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Predicted probability of live birth

C


Observed proportion with live birth
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Predicted probability of live birth

D

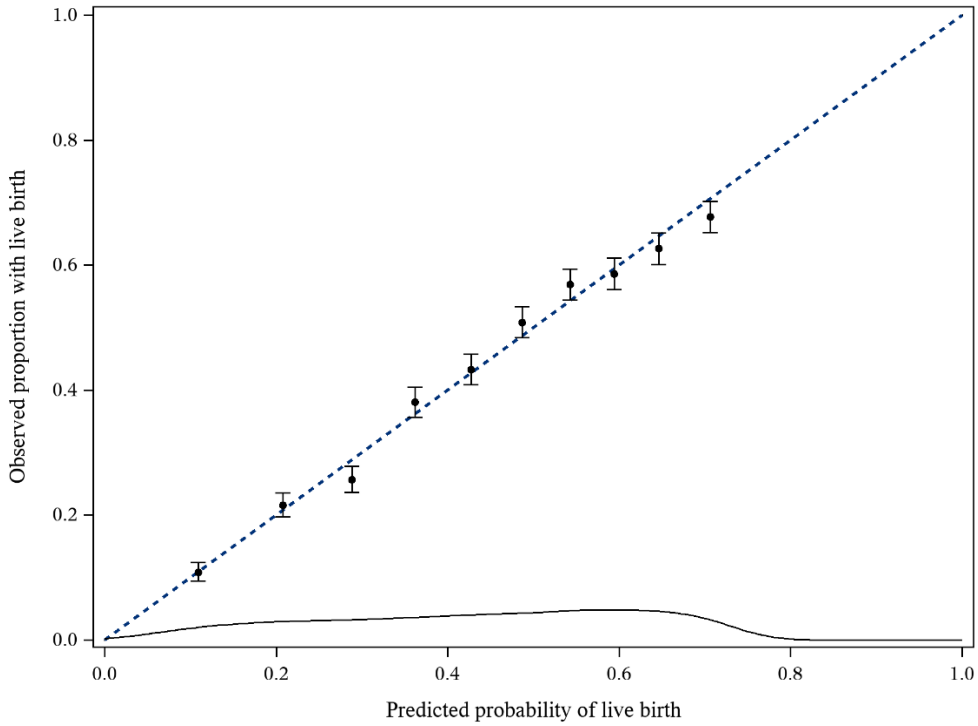


Figure 2 Plots showing the predicted probability of live birth over the range of AMH (ng/ml) for women of different ages with unexplained infertility and BMI of 24

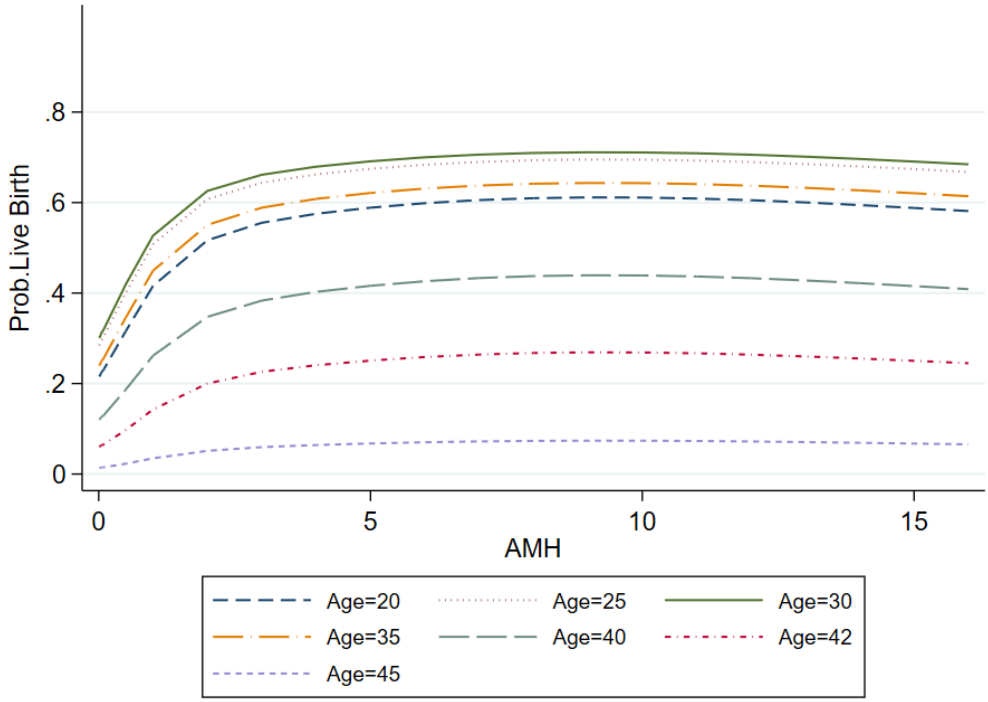


Table 1 Characteristics of couples and their treatment before the first complete cycle and before the second complete cycle of in vitro fertilization (IVF), values are numbers (%) unless stated otherwise

Characteristics	Before first complete cycle		Before second complete cycle
	All eligible women	Women with AMH	All eligible women
No of women	88,613	53,766	24,735
No of complete cycles (up to max of 3)	121,561	73,603	32,948
Complete cycles			
1	88,613 (100)	53,766 (100)	24,735 (100)
2	24,735 (20.3)	14,945 (20.3)	8,213 (24.9)
3	8,213 (6.8)	4,892 (6.6)	
Patients characteristics			
Age of women (yrs), mean (SD)	34.6 (4.8)	34.4 (4.7)	36.5 (4.7)
Previous Full-Term Birth			
No	71,499 (80.7)	43,605 (81.1)	20,190 (81.6)
Yes	17,114 (19.3)	10,161 (18.9)	4,545 (18.4)
Missing	333 (0.4)	108 (0.2)	68 (0.2)
Cause of infertility			
Male factor	32,481 (36.7)	21,123 (39.3)	8,580 (34.7)
Tubal factor	13,873 (15.7)	8,572 (15.9)	3,508 (14.2)
Diminished ovarian reserve	21,740 (24.5)	13,446 (25.0)	9,616 (38.9)
Polycystic ovaries or PCOS	24,294 (27.4)	15,317 (28.5)	9,448 (38.2)
Endometriosis	7,286 (8.2)	4,583 (8.5)	1,984 (8.0)
Uterine factor	4,362 (4.9)	2,797 (5.2)	1,396 (5.6)
Unexplained	13,934 (15.7)	7,712 (14.3)	3,396 (13.7)
BMI (Kg/m^2), median (IQR)	24.4 (21.7, 28.7)	24.4 (21.7, 28.8)	24.6 (21.8, 29.1)
Missing	18,246 (20.6)	8,511 (15.8)	4,976 (20.1)
AMH (ng/ml), median (IQR)	2.4 (1.1, 4.4)	2.4 (1.1, 4.4)	1.3 (0.6, 2.7)
Missing	34,847 (39.3)	-	13,875 (56.1)
Eggs collected at the first complete cycle, median (IQR)			7 (2, 12)

Table 2 Adjusted logistic regression models predicting cumulative live birth using predictors available before the first complete IVF cycle (pre-treatment models) and before the second complete IVF cycle (post-treatment model)

Predictors	Pre-treatment model		Post-treatment model
	Adjusted model for all women OR (95% CI)	Adjusted model for women with AMH OR (95% CI)	Adjusted model for all women OR (95% CI)
	N=88,613	N=53,766	N=24,735
Complete cycle number			
1	1	1	
2	0.58 (0.56, 0.60)	0.61 (0.58, 0.63)	1
3	0.41 (0.38, 0.43)	0.45 (0.42, 0.49)	0.69 (0.65, 0.74)
Woman's Age (yrs)[a]			
20	0.74 (0.67, 0.82)	0.69 (0.60, 0.79)	0.88 (0.72, 1.07)
Reference=25	1	1	1
30	1.03 (0.98, 1.09)	1.09 (1.02, 1.17)	1.01 (0.89, 1.15)
35	0.69 (0.66, 0.73)	0.79 (0.74, 0.85)	0.80 (0.69, 0.92)
40	0.28 (0.26, 0.30)	0.34 (0.32, 0.37)	0.39 (0.34, 0.45)
45	0.04 (0.03, 0.04)	0.03 (0.03, 0.04)	0.09 (0.07, 0.12)
Previous Full-Term Birth			
No	1	1	-
Yes	1.05 (1.01, 1.08)	1.05 (1.01, 1.10)	-
Type of infertility (yes v no)			
Male factor	1.17 (1.14, 1.20)	1.08 (1.04, 1.12)	1.18 (1.11, 1.24)
Polycystic ovary syndrome	1.15 (1.12, 1.19)	1.04 (0.99, 1.08)	1.14 (1.07, 1.22)
Uterine factor	0.82 (0.77, 0.87)	0.84 (0.78, 0.90)	0.75 (0.67, 0.85)
Diminished ovarian reserve	0.51 (0.50, 0.53)	0.84 (0.80, 0.89)	0.66 (0.61, 0.71)
Unexplained	1.10 (1.05, 1.14)	1.07 (1.01, 1.12)	-
Woman's BMI (Kg/m^2)[a]			
19	1.02 (0.98, 1.06)	0.99 (0.93, 1.04)	0.91 (0.84, 0.99)
23	1.06 (1.04, 1.08)	1.05 (1.03, 1.08)	1.03 (0.99, 1.08)
Reference=25	1.00	1.00	1.00
30	0.86 (0.84, 0.89)	0.87 (0.83, 0.90)	0.89 (0.84, 0.95)
35	0.75 (0.72, 0.78)	0.76 (0.72, 0.80)	0.80 (0.73, 0.87)
40	0.66 (0.63, 0.69)	0.68 (0.63, 0.72)	0.72 (0.65, 0.80)
AMH (ng/ml) [a]			
1		0.61 (0.58, 0.64)	
2		0.91 (0.90, 0.93)	
Reference=2.5		1	
5		1.22 (1.18, 1.27)	
10		1.34 (1.26, 1.43)	
15		1.22 (1.11, 1.34)	
Number of eggs collected at the first complete cycle [a]			
5			0.76 (0.71, 0.81)
Reference=9			1
15			1.10 (1.03, 1.18)
20			1.09 (1.00, 1.18)

25			1.02 (0.93, 1.12)

[a]Age, BMI, AMH and number of eggs collected had a non-linear relationship with the log odds of live birth and so were fitted to the model as restricted cubic splines with 5 knots. To aid interpretation we have provided odds ratios for different age, BMI, AMH and egg number values along these spline curves versus a given reference value.

Supplementary Appendix

Prediction of personalized cumulative live birth before the first and second complete cycle of in-vitro fertilisation (IVF): a population-based study of linked cycle data from 79,512 women

David J McLernon, Amalraj Raja, James P Toner, Valerie L. Baker, Kevin Doody, David B Seifer, Amy E Sparks, Ethan Wantman, Paul Lin, Siladitya Bhattacharya, Bradley Van Voorhis

Table of contents

Text S1. Further detail on missing data analysis

Table S1 shows the characteristics of women with BMI and women without BMI recorded in SART-CORS. The distribution of characteristics was very similar between these two groups. Single imputation was performed to impute missing BMI values. The imputation was performed using the Markov Chain Monte Carlo method. We used all the baseline characteristics described in the paper, last complete cycle number, and live-birth outcome to inform the imputation process.[1]

Single imputation was used to impute missing BMI values and previous full-term live birth status in 21% of couples in order to increase statistical power and adjust for any biases caused by excluding women with missing BMI. Single rather than multiple imputation was used because methods to assess model performance such as discrimination and calibration are much more straightforward to calculate using the former. The database was very large, and the amount of missing data was relatively low meaning that there was minimal risk of underestimation of the uncertainty associated with imputed values as can arise when using single imputation in small datasets.[2]

Text S2. Formulas for all SART prediction models

About this tool

This IVF prediction tool was developed using patient and treatment data from the Society of Assisted Reproductive Technology (SART) who collect and store data on all licensed fertility treatments centres in the USA. This tool is for the use of heterosexual couples who are starting IVF treatment for the first time using their own eggs and sperm.

Before you begin your treatment, SART IVF Pre-treatment calculates your chance of having a baby over one or more complete cycles of IVF. A complete cycle means all fresh and frozen-thawed embryo transfers resulting from one episode of ovarian stimulation.

Once you have had your first embryo transfer, SART IVF Post-Treatment uses this information to update your chance of having a baby.

The SART IVF calculators were developed using patient and treatment data from all IVF cycles carried out in the USA from 2014 to 2016. All of the personal identifying information was removed from the data before being used in the creation of the tool.

It is important to note that your predicted chances of having a baby calculated from the SART tool are not necessarily accurate. There are many other unmeasurable characteristics of patients which were not available when developing the tool that may improve or decrease your individual chance of having a baby. If you are concerned about your predicted chance, then please consult your family physician.

SART Pre-treatment Models

Predicts the cumulative probability of live birth over three complete cycles from just before the start of the first complete cycle

Below are the Predictors with their description for webpage (and range of possible values)

Age - What age are you? (18 to 50 years)

Body Mass Index (BMI) – What is your body mass index? (16 to 50)

Previous Full-Term birth - Have you ever had a baby born to full term i.e. over 37 weeks? (0=No; 1=Yes)

Male Factor – Does your partner have a problem with their sperm? (1=Yes; 0=No)

Polycystic disorder – Do you have polycystic ovaries or polycystic ovary syndrome? (1=Yes; 0=No)

Uterine – Do you have any of the following uterine problems: septum, myoma, intrauterine adhesions, congenital anomalies? (1=Yes; 0=No)

Unexplained – Do you have unexplained infertility? (1=Yes; 0=No)

Diminished ovarian reserve – Have you been diagnosed as having a low ovarian reserve? (1=Yes; 0=No)

Do you know your most recent AMH level? YES or NO

If the respondent answers NO, then provide predictions from MODEL 1 below

If the respondent answers YES, then they should enter their AMH in ng/ml, and provide predictions from MODEL 2 below. The value of AMH ranges from 0.01 to 16 in the model. However, if someone enters more than 16 for AMH then set to 16 for calculation.

Pre-treatment model formula (excluding serum AMH):

1. The following Age1 to Age4 equations are first calculated using the female age and then feed into the XB equation below in point 3:

Age1 = age-25
Age2 = (max(age-32,0)^3) – (0.5625 * (max(age-18, 0)^3)) – (0.4375 * (max(age-50, 0)^3)) + 192.9375
Age3 = (max(age-35, 0)^3) – (0.46875 * (max(age-18, 0)^3)) – (0.53125 * (max(age-50, 0)^3)) + 160.78125
Age4 = (max(age-39,0)^3) – (0.34375 * (max(age-18, 0)^3)) – (0.65625 * (max(age-50, 0)^3)) + 117.90625

where ^ means 'to the power of'

2. Please calculate the following bmi1 to bmi4 equations using the female bmi and then feed into the XB equation below in point 3:

bmi1 = bmi-25
bmi2 = (max(bmi-21.7,0)^3) – (0.83235294 * (max(bmi-16, 0)^3)) – (0.16764706 * (max(bmi-50, 0)^3)) +570.8483
bmi3 = (max(bmi-24.3, 0)^3) – (0.75588238 * (max(bmi-16, 0)^3)) – (0.24411762 * (max(bmi-50, 0)^3)) +550.6953
bmi4 = (max(bmi-28.8,0)^3) – (0.62352943 * (max(bmi-16, 0)^3)) – (0.37647057 * (max(bmi-50, 0)^3)) + 454.55295

where ^ means 'to the power of'

3. Calculate XB

XB= 0.4152001
+ (0.0767695 * Age1) + (0.0010931 * Age2) + (-0.0029217 * Age3) + (0.0029371 * Age4)
+ (0.0610799 * bmi1) + (0.0021156 * bmi2) + (-0.0014594 * bmi3) + (-0.00000542 * bmi4)
+ (0.0444797 * FullTermBirths) + (0.157452 * MaleInfertility)
+ (0.1418316 * polycpcos) + (-0.1979537 * Uterine)
+ (0.0915996 * Unexplained) + (-0.6658235 * OvulDisorder)

4. For each couple we want to calculate their probability of live birth after the first, second, and third cycle of IVF:

PCycle1 = exp(XB)/(1+exp(XB))

PCycle2 = exp(XB - 0.5453622)/(1+exp(XB - 0.5453622))

PCycle3 = exp(XB - 0.898559)/(1+exp(XB - 0.898559))

5. We then calculate the cumulative probability of a live birth after 1, 2, and 3 cycles:

CumPCycle1 = 1-(1- PCycle1)

CumPCycle2 = 1-((1- PCycle1)*(1- PCycle2))

CumPCycle3 = 1-((1- PCycle1)*(1- PCycle2)*(1- PCycle3))

Pre-treatment model formula (including serum AMH):

1. The following Age1 to Age4 equations are first calculated using the female age and then feed into the XB equation below in point 4:

Age1 = age-25
Age2 = (max(age-31,0)^3) – (0.59375 * (max(age-18, 0)^3)) – (0.406250 * (max(age-50, 0)^3)) + 203.6563
Age3 = (max(age-35, 0)^3) – (0.46875 * (max(age-18, 0)^3)) – (0.53125 * (max(age-50, 0)^3)) + 160.78125
Age4 = (max(age-39,0)^3) – (0.34375 * (max(age-18, 0)^3)) – (0.65625 * (max(age-50, 0)^3)) + 117.90625

where ^ means 'to the power of'

2. Please calculate the following bmi1 to bmi4 equations using the female BMI and then feed into the XB equation below in point 4:

bmi1 = bmi-25
bmi2 = (max(bmi-21.7,0)^3) – (0.83235294 * (max(bmi-16, 0)^3)) – (0.16764706 * (max(bmi-50, 0)^3)) +570.84833
bmi3 =(max(bmi-24.399, 0)^3) – (0.75294119 * (max(bmi-16, 0)^3)) – (0.24705881 * (max(bmi-50, 0)^3)) +548.6781
bmi4 = (max(bmi-28.799,0)^3) – (0.62352943 * (max(bmi-16, 0)^3)) – (0.37647057 * (max(bmi-50, 0)^3)) + 454.55295

where ^ means 'to the power of'

3. Please calculate the following amh1 to amh4 equations using the AMH measurement and then feed into the XB equation below in point 4:

amh1=amh-2.5
amh2=(max(amh-0.98,0)^3) – (0.93933707 * max(amh-0.01, 0)^3) – (0.06066293 * max(amh-16, 0)^3) + 10.989912
amh3=(max(amh-2.0999, 0)^3) - (0.86929959 * max(amh-0.01, 0)^3) – (0.13070041 * max(amh-16, 0)^3) + 13.356415

amh4=(max(amh-4.03,0)^3) - (0.74859285 * max(amh-0.01, 0)^3) – (0.25140715 * max(amh-16, 0)^3) + 11.556963

where ^ means 'to the power of'

4. Calculate XB

XB= 0.4346214
+ (0.0920238 * (Age1)) + (0.0011043 * (Age2)) + (-0.0039663 * (Age3)) + (0.0042808 * (Age4))
+ (0.0664307 * (bmi1)) + (0.0019531 * (bmi2)) + (-0.0012456 * (bmi3)) + (-0.0000662 * (bmi4))
+ (1.095414 * (amh1)) + (0.234447 * (amh2)) + (-0.0890884 * (amh3)) + (-0.009418 * (amh4))
+ (0.0495487 * (FullTermBirths)) + (0.0803214 * (MaleInfertility))
+ (0.0349664 * (polycpcos)) + (-0.1772406 * (Uterine)) + (0.0656224 * (Unexplained))
+ (-0.1695679 * (OvulDisorder))

5. For each couple we want to calculate their probability of live birth after the first, second, and third cycle of IVF:

PCycle1 = exp(XB)/(1+exp(XB))

PCycle2 = exp(XB - 0.4993235)/(1+exp(XB - 0.4993235))

PCycle3 = exp(XB - 0.7894117)/(1+exp(XB - 0.7894117))

6. We then calculate the cumulative probability of a live birth after 1, 2, and 3 cycles:

CumPCycle1 = 1-(1- PCycle1)

CumPCycle2 = 1-((1- PCycle1)*(1- PCycle2))

CumPCycle3 = 1-((1- PCycle1)*(1- PCycle2)*(1- PCycle3))

SART Post-treatment model

Predicts the cumulative probability of live birth over two complete cycles from just before the start of the second complete cycle

Below are the Predictors with their description for webpage (and range of possible values)

Age - What age are you? (18 to 50 years)

Body Mass Index (BMI) – What is your body mass index? (16 to 50)

Male Factor – Does your partner have a problem with their sperm? (1=Yes; 0=No)

Polycystic disorder – Do you have polycystic ovaries or polycystic ovary syndrome? (1=Yes; 0=No)

Uterine – Do you have any of the following uterine problems: septum, myoma, intrauterine adhesions, congenital anomalies? (1=Yes; 0=No)

Diminished ovarian reserve – Have you been diagnosed as having a low ovarian reserve? (1=Yes; 0=No)

Eggs – How many eggs were collected from your first IVF cycle? (0 to 29, any number above 29 is coded as 29)

Post-treatment model formula:

1. The following Age1 to Age4 equations are first calculated using the female age and then feed into the XB equation below in point 4:

Age1 = age-25
Age2 = (max(age-33,0)^3) – (0.5666666 * (max(age-20, 0)^3)) – (0.43333334 * (max(age-50, 0)^3)) + 70.833336
Age3 = (max(age-37, 0)^3) – (0.43333334 * (max(age-20, 0)^3)) – (0.566666 * (max(age-50, 0)^3)) + 54.166668
Age4 = (max(age-40,0)^3) – (0.33333334 * (max(age-20, 0)^3)) – (0.6666666 * (max(age-50, 0)^3)) + 41.666668
where ^ means 'to the power of'

2. Please calculate the following bmi1 to bmi4 equations using the female BMI and then feed into the XB equation below in point 4:

bmi1 = bmi-25
bmi2 = (max(bmi-21.7,0)^3) – (0.83679527 * (max(bmi-16.2, 0)^3)) – (0.16320473 * (max(bmi-49.9, 0)^3)) +534.31543
bmi3 = (max(bmi-24.5, 0)^3) – (0.75370926 * (max(bmi-16.2, 0)^3)) – (0.24629074 * (max(bmi-49.9, 0)^3)) +513.50659
bmi4 = (max(bmi-29,0)^3) – (0.62017804 * (max(bmi-16.2, 0)^3)) – (0.37982196 * (max(bmi-49.9, 0)^3)) + 422.63388

where ^ means 'to the power of'

3. Please calculate the following Retr1 to Retr4 equations using the number of Eggs collected and then feed into the XB equation below in point 4:

gen Retr1=Retr-9
gen Retr2=(max(Retr-2,0)^3) – (0.93103451 * (max(Retr-0, 0)^3)) – (0.06896549 * (max(Retr-29, 0)^3)) + 335.72415
gen Retr3=(max(Retr-7, 0)^3) - (0.75862068 * (max(Retr-0, 0)^3)) – (0.24137932 * (max(Retr-29, 0)^3)) + 545.03448
gen Retr4=(max(Retr-13,0)^3) - (0.58620691 * (max(Retr-0, 0)^3)) – (0.41379309 * (max(Retr-29, 0)^3)) + 427.34482

where ^ means 'to the power of'

4. Calculate XB

XB = -0.1404629
+ (0.0307055 * (Age1)) + (-0.000268 * (Age2)) + (-0.0003146 * (Age3)) + (0.0013544 * (Age4))
+ (0.0828801 * (bmi1)) + (0.0019979 * (bmi2)) + (-0.0011626 * (bmi3)) + (-0.0001167 * (bmi4))
+ (-0.0697645 * (Retr1)) + (-0.0072497 * (Retr2)) + (0.0031553 * (Retr3)) + (-0.0005503 * (Retr4)) + (0.1618111 * (MaleInfertility)) + (-0.4128434 * (OvulDisorder)) + (0.1345618 * (polycpcos)) + (-0.285821 * (Uterine))

5. For each couple we want to calculate their probability of live birth after the first, second, and third cycle of IVF:

PCycle2 = exp(XB)/(1+exp(XB))

PCycle3 = exp(XB -0.3670816)/(1+exp(XB -0.3670816))

6. We then calculate the cumulative probability of a live birth after 2, and 3 cycles:

CumPCycle2 = 1-(1- PCycle2)

CumPCycle3 = 1-((1- PCycle2)*(1- PCycle3))

A sentence with plot appears saying 'Your chance of having a baby after ? cycle(s) of IVF is: X%. This means that out of 100 couples having ? cycles of IVF, approximately X would have a baby'.

Figure S1 Flow Chart of exclusion criteria

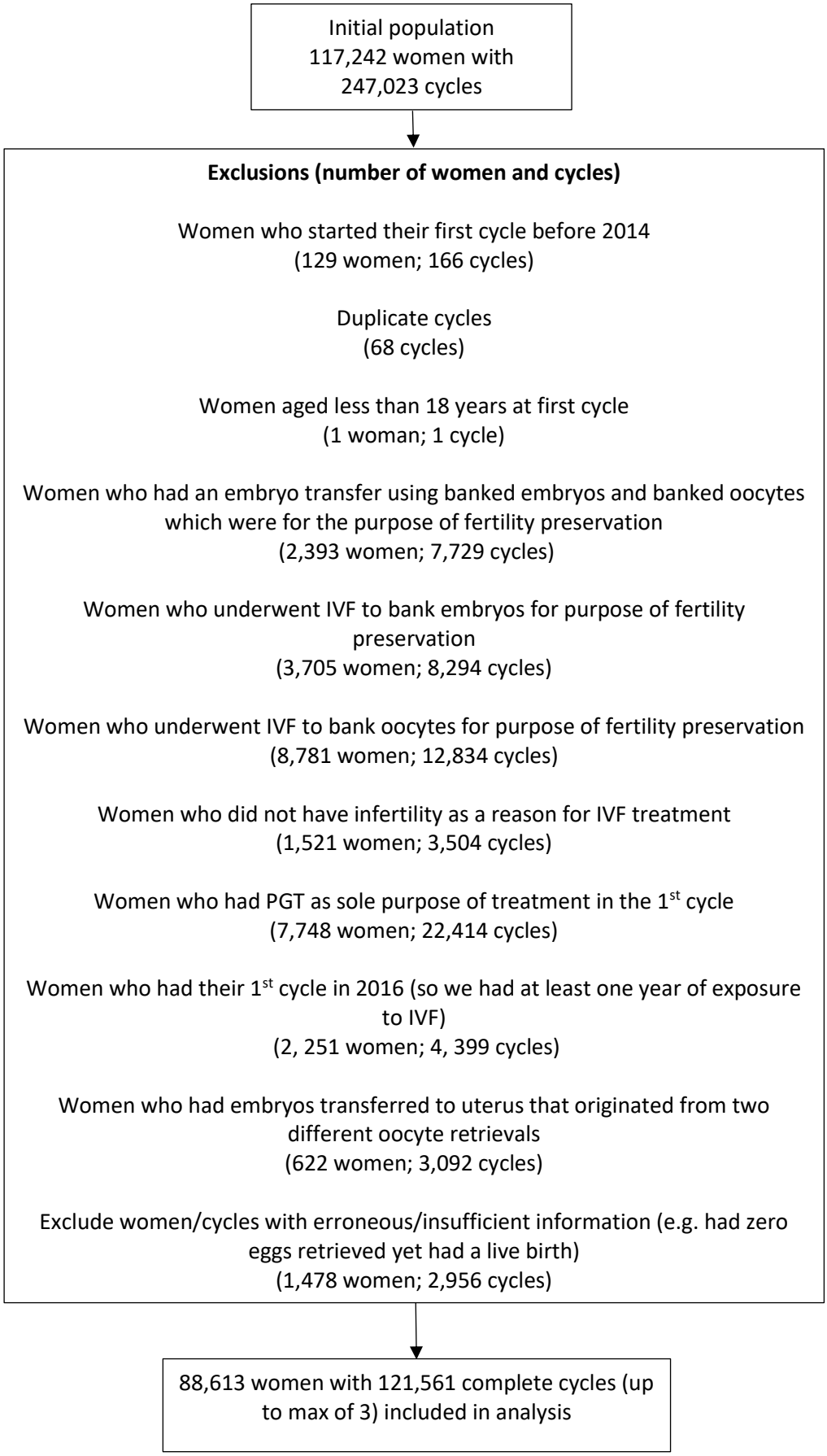


Figure S2 Number of women having a live birth, not having a live birth but continuing treatment, and not having a live birth and not continuing treatment over three complete cycles of IVF

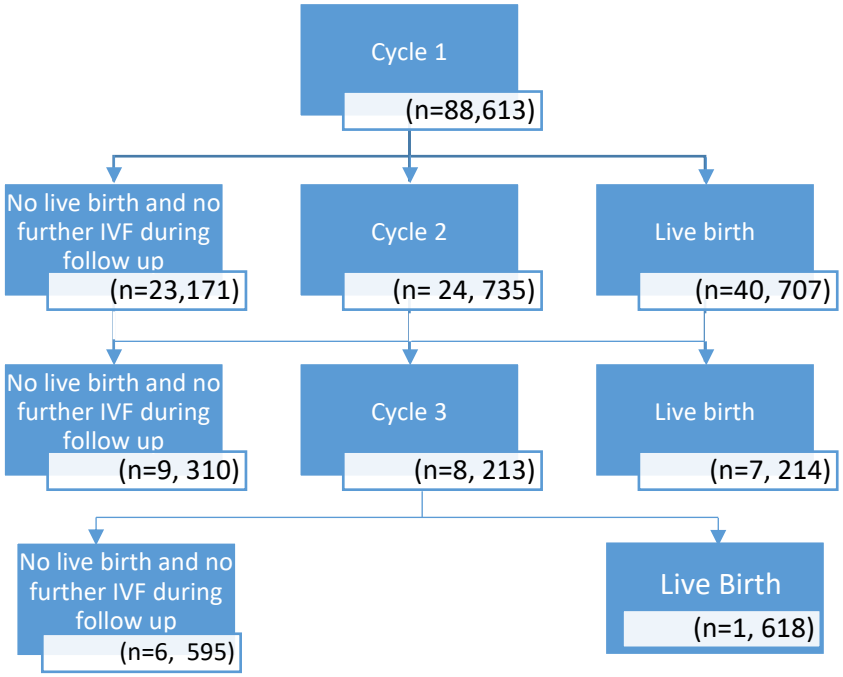
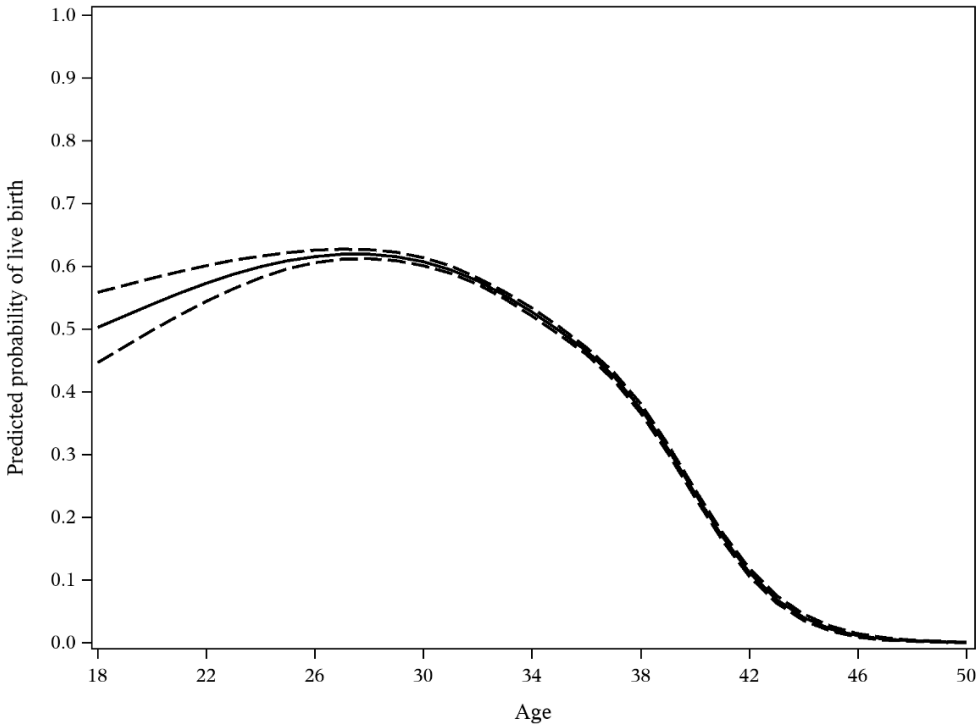
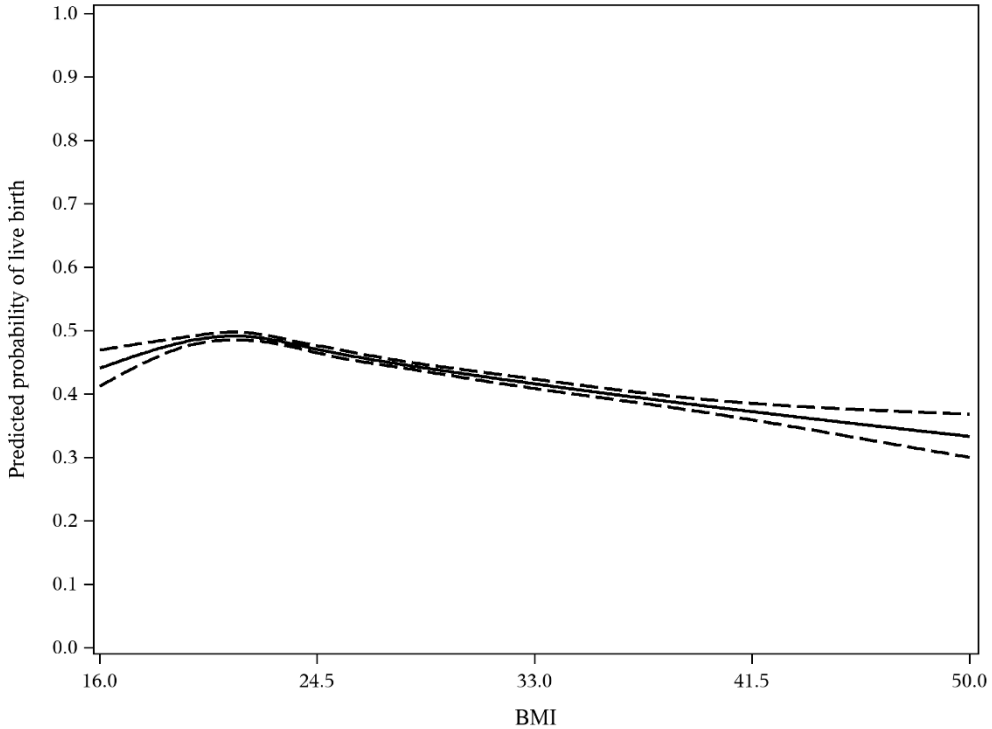


Figure S3 Plots showing unadjusted (univariate) relationship between A woman's age (years); B Body Mass Index; C AMH (ng/ml) and live birth after one complete cycle of IVF, and between D Number of eggs collected in the first complete cycle and live birth after two complete cycles of IVF

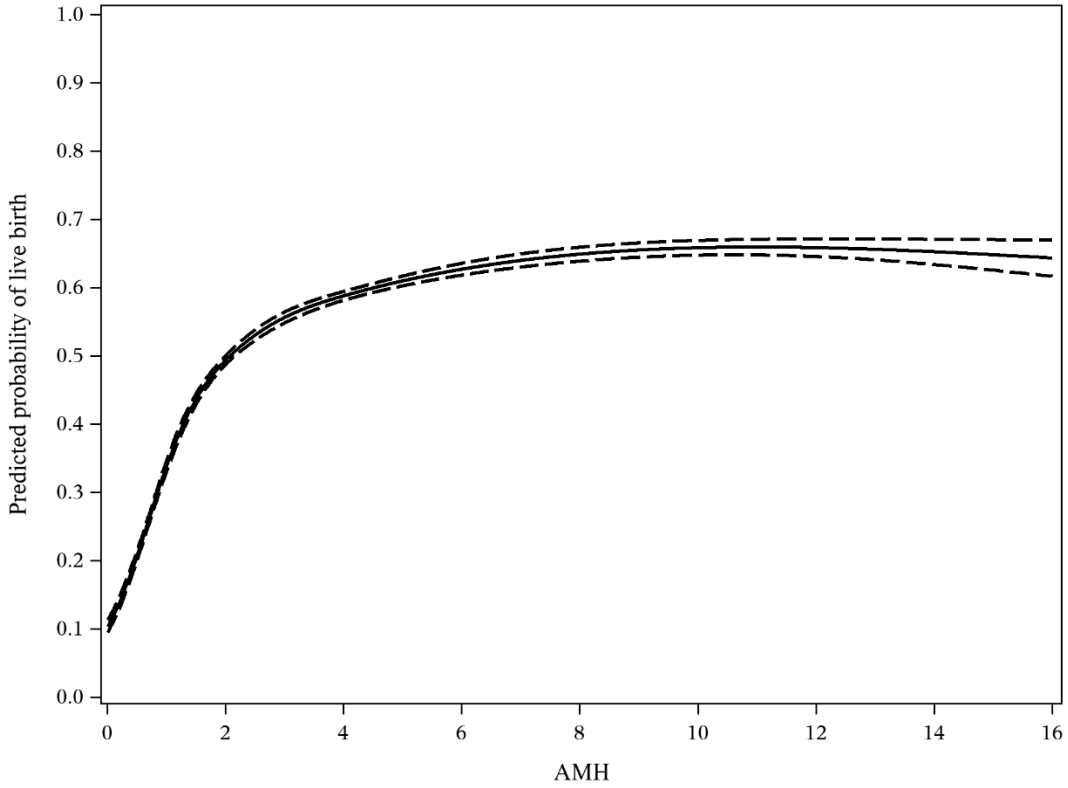
A



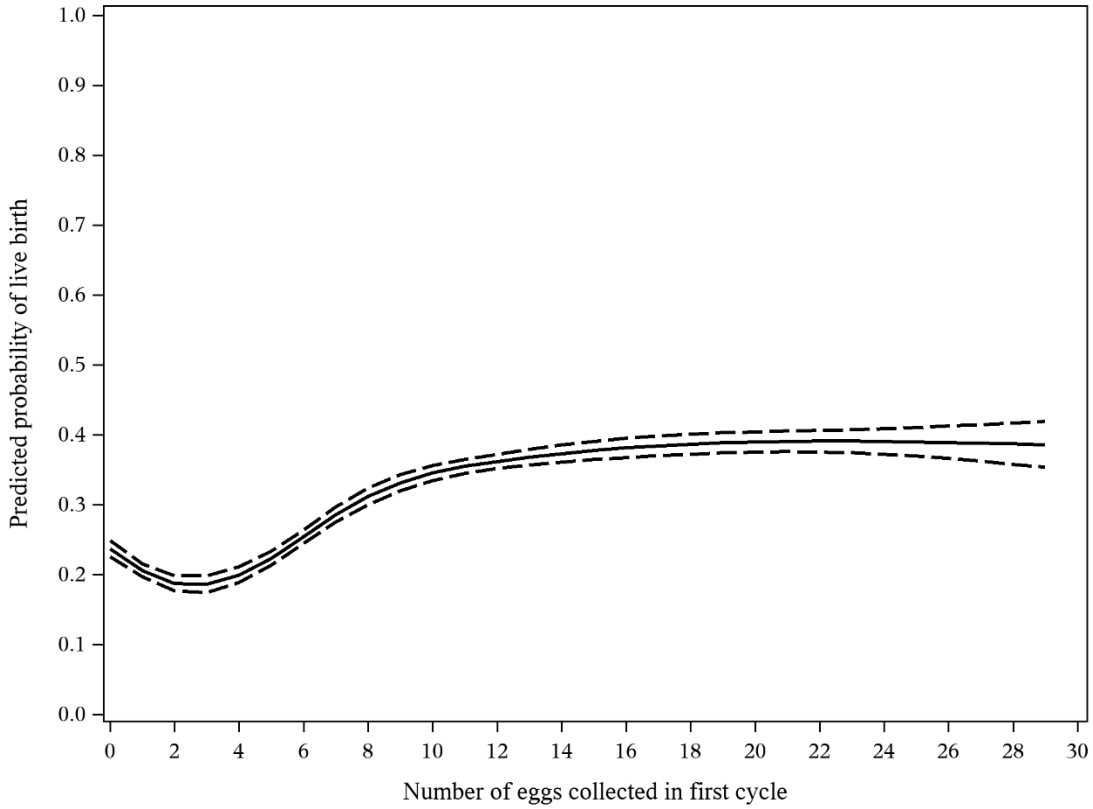
B



C



D



Note: Each panel depicts the probability of live-birth (solid curve) with 95% confidence bands as a function of the baseline variable.

Table S1 Characteristics of women who had missing BMI and women who did not have a missing BMI, n (%) unless otherwise stated

Characteristics	Women with a BMI measurement	Women with missing BMI measurement
No of women	70,367	18,246
Complete cycles		
1	70,367 (100.0)	18,246 (100.0)
2	19,759 (20.4)	4,976 (20.0)
3	6,537 (6.8)	1,676 (6.7)
Patients characteristics		
Age of women (yrs), mean (SD)	34.5 (4.8)	34.7 (4.8)
Type of infertility		
Primary infertility	40,580 (57.7)	10,786 (59.1)
Secondary infertility	29,787 (42.3)	7,460 (40.9)
Previous Full-Term Birth		
No	56,793 (80.7)	14,706 (80.6)
Yes	13,574 (19.3)	3,540 (19.4)
Cause of infertility		
Male factor	26,397 (37.5)	6,084 (33.3)
Tubal factor	11,232 (16.0)	2,641 (14.5)
Diminished ovarian reserve	16,680 (23.7)	5,060 (27.7)
Polycystic ovaries or PCOS	19,182 (27.3)	5,112 (28.0)
Endometriosis	5,943 (8.4)	1,343 (7.4)
Uterine factor	3,578 (5.1)	784 (4.3)
Unexplained	11,060 (15.7)	2,874 (15.8)
AMH (ng/ml) (n=49,727), median (IQR)	2.4 (1.1, 4.4)	2.4 (1.1, 4.5)
missing	26,219 (36.7)	11,173 (56.8)
Eggs retrieved at first cycle, median (IQR)	11 (6, 17)	11 (6,17)

Table S2 Performance statistics for all prediction models

Performance statistic	Pre-treatment model		Post-treatment model
	No AMH	AMH	
Cumulative live birth rate, n (%)	49,539 (55.9%)	30,839 (57.4%)	8,832 (35.7%)
McFadden R-squared	12.4%	13.9%	8.6%
C-statistic	0.71	0.73	0.71
Heuristic shrinkage	1.00	1.00	0.99

References

1. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Br Med J 2009; 338: b2393.

2. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer, 2009.