Human Reproduction, pp. 1–12, 2022
https://doi.org/10.1093/humrep/deac152

human reproduction

ORIGINAL ARTICLE *Infertility*

Predicting cumulative live birth for couples beginning their second complete cycle of *in vitro* fertilization treatment

Mariam B. Ratna [1,2,*], Siladitya Bhattacharya[3], N. van Geloven[4], and David J. McLernon [1]

[1]Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK [2]Warwick Clinical Trial Units, Warwick Medical School, University of Warwick, Coventry, UK [3]School of Medicine, Medical Sciences & Nutrition, University of Aberdeen, Aberdeen, UK [4]Department of Biomedical Data Sciences, Medical Statistics, Leiden University Medical Center, Leiden, The Netherlands

*Correspondence address. Institute of Applied Health Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK. E-mail: s06mr1@abdn.ac.uk, mariam.ratna@warwick.ac.uk https://orcid.org/0000-0001-9468-6688

Submitted on December 9, 2021; resubmitted on June 1, 2022; editorial decision on June 20, 2022

STUDY QUESTION: Can we develop an IVF prediction model to estimate individualized chances of a live birth over multiple complete cycles of IVF in couples embarking on their second complete cycle of treatment?

SUMMARY ANSWER: Yes, our prediction model can estimate individualized chances of cumulative live birth over three additional complete cycles of IVF.

WHAT IS KNOWN ALREADY: After the completion of a first complete cycle of IVF, couples who are unsuccessful may choose to undergo further treatment to have their first child, while those who have had a live birth may decide to have more children. Existing prediction models can estimate the overall chances of success in couples before commencing IVF but are unable to revise these chances on the basis of the couple's response to a first treatment cycle in terms of the number of eggs retrieved and pregnancy outcome. This makes it difficult for couples to plan and prepare emotionally and financially for the next step in their treatment.

STUDY DESIGN, SIZE, DURATION: For model development, a population-based cohort was used of 49 314 women who started their second cycle of IVF including ICSI in the UK from 1999 to 2008 using their own oocytes and their partners' sperm. External validation was performed on data from 39 442 women who underwent their second cycle from 2010 to 2016.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Data about all UK IVF treatments were obtained from the Human Fertilisation and Embryology Authority (HFEA) database. Using a discrete time logistic regression model, we predicted the cumulative probability of live birth from the second up to and including the fourth complete cycles of IVF. Inverse probability weighting was used to account for treatment discontinuation. Discrimination was assessed using c-statistic and calibration was assessed using calibration-in-the-large and calibration slope.

MAIN RESULTS AND THE ROLE OF CHANCE: Following exclusions, 49 314 women with 73 053 complete cycles were included. 12 408 (25.2%) had a live birth resulting from their second complete cycle. Cumulatively, 17 394 (35.3%) had a live birth over complete cycles two to four. The model showed moderate discriminative ability (c-statistic: 0.65, 95% CI: 0.64 to 0.65) and evidence of overprediction (calibration-in-the-large = −0.08) and overfitting (calibration slope 0.85, 95% CI: 0.81 to 0.88) in the validation cohort. However, after recalibration the fit was much improved. The recalibrated model identified the following key predictors of live birth: female age (38 versus 32 years—adjusted odds ratio: 0.59, 95% CI: 0.57 to 0.62), number of eggs retrieved in the first complete cycle (12 versus 4 eggs; 1.34, 1.30 to 1.37) and outcome of the first complete cycle (live birth versus no pregnancy; 1.78, 1.66 to 1.91; live birth versus pregnancy loss; 1.29, 1.23 to 1.36). As an example, a 32-year-old with 2 years of non-tubal infertility who had 12 eggs retrieved from her first stimulation and had a live birth during her first complete cycle has a 46% chance of having a further live birth from the second complete cycle of IVF and an 81% chance over a further three cycles.

LIMITATIONS, REASONS FOR CAUTION: The developed model was updated using validation data that was 6 to 12 years old. IVF practice continues to evolve over time, which may affect the accuracy of predictions from the model. We were unable to adjust for some

potentially important predictors, e.g. BMI, smoking and alcohol intake in women, as well as measures of ovarian reserve such as antral follicle count. These were not available in the linked HFEA dataset.

WIDER IMPLICATIONS OF THE FINDINGS: By appropriately adjusting for couples who discontinue treatment, our novel prediction model will provide more realistic chances of live birth in couples starting a second complete cycle of IVF. Clinicians can use these predictions to inform discussion with couples who wish to plan ahead. This prediction tool will enable couples to prepare emotionally, financially and logistically for IVF treatment.

STUDY FUNDING/COMPETING INTEREST(S): This study was supported by an Elphinstone scholarship scheme at the University of Aberdeen and Aberdeen Fertility Centre, University of Aberdeen. The authors have no conflict of interest.

TRIAL REGISTRATION NUMBER: N/A.

Key words: cumulative livebirth / IVF / prediction model / treatment discontinuation / Inverse Probability Weighting

Introduction

Over the last three decades, more than 30 clinical prediction models have been developed to estimate individualized chances of IVF success (Wiegerinck *et al.*, 1999; Van Der Steeg *et al.*, 2006; Ratna *et al.*, 2020). Conventionally reported as live birth rate in a single fresh cycle (Abuzeid *et al.*, 2014; Vrtacnik *et al.*, 2014), the widespread use of embryo cryopreservation in the past two decades has meant that cumulative live birth rates (CLBRs), including outcomes following fresh as well as frozen embryo transfers are more informative and clinically relevant (Maheshwari *et al.*, 2015; McLernon *et al.*, 2016a). Two recent studies have estimated the cumulative probability of live birth per woman following one or more treatment cycles in couples starting IVF (Luke *et al.*, 2014; McLernon *et al.*, 2016b). Luke *et al.* (2014) estimated the cumulative chance of live birth per woman over three fresh treatment cycles but excluded the contribution of any subsequent frozen embryo transfers. The McLernon (pre- and post-treatment) models predict the cumulative probability of live birth per woman across successive complete cycles of IVF, where a complete cycle includes all fresh and frozen-thawed embryo transfers resulting from a single episode of ovarian stimulation. The pre-treatment model provides cumulative predictions before the start of the first complete cycle, and the post-treatment model updates these predictions after the first fresh embryo transfer attempt but before the transfer of all available frozen embryos (McLernon *et al.*, 2016b).

After an unsuccessful first complete cycle of IVF, about 40% of couples embark on a second cycle in order to have their first child (McLernon *et al.*, 2016b). Those who have had a live birth may decide to have more children. At this stage, key clinical characteristics, such as age and duration of infertility will be different from those at the start of the first cycle and these along with a couple's response to the first complete cycle are likely to influence chances of live birth in the second and subsequent attempts (Malizia *et al.*, 2013). As some couples discontinue treatment, the change in the nature of the patient cohort over successive complete cycles needs to be taken into account for an accurate prediction of success.

None of the current prediction models are able to predict IVF live birth for all women starting a second complete cycle. This is an unmet need for many couples who want to plan and prepare emotionally and financially for the next step in their treatment. In this study, we aimed to derive and validate a novel prediction model that can estimate the overall chances of a live birth in couples who have completed an IVF cycle and are about to start a second complete cycle.

Materials and methods

Database

Since 1992, all licenced IVF treatments in the UK have been collected by the Human Fertilisation and Embryology Authority (HFEA). A detailed version of the HFEA database contains data on all treatment cycles linked to women undergoing IVF treatment, which enables researchers to estimate the cumulative probability of live birth per woman. An anonymized version of this linked database was obtained following approval from the North of Scotland research ethics committee, the Confidentiality Advisory Group and the HFEA register research panel.

Study population

This population-based cohort study used linked data from 49 314 women who underwent ovarian stimulation for a second complete cycle of IVF in the UK between January 1999 and September 2008. Data on frozen embryo transfer episodes were extracted up until September 2009 so that all women had a follow-up period of at least 1 year after their fresh embryo transfer. A second cohort, used for external validation of the model, comprised IVF records from 39 442 women who had ovarian stimulation for a second complete cycle between January 2010 and December 2016. Frozen embryo transfer episodes connected to these stimulation episodes were included up until December 2017. Only women whose treatment involved using their own oocytes and/or their partners' sperm were included in the study sample, i.e. women whose treatment involved donor insemination, egg donation or surrogacy were excluded. Women whose treatment cycles were cancelled prior to eggs retrieval due to poor ovarian response were also included in the study. If this occurred in the first complete cycle, then the outcome was classified as 'no pregnancy'.

Baseline characteristics

Couple and treatment characteristics at the commencement of the second complete cycle, selected as the baseline parameters, included woman's age (years), duration of infertility (years), time since the first egg retrieval (defined as difference in the date of egg retrieval between the first and second complete cycles) (months), causes of infertility (tubal, male factor, anovulation, unexplained or endometriosis), year of egg retrieval and type of treatment (ICSI or IVF). The causes of infertility were not mutually exclusive. For example, a woman diagnosed with tubal and ovulatory infertility would have both causes.

Information on response to the first complete cycle included number of eggs retrieved and pregnancy outcome (no pregnancy, live birth, pregnancy loss). Women who had an ectopic pregnancy, biochemical pregnancy, miscarriage, molar pregnancy, termination, still birth or embryo reduction following the first complete cycle were included in the 'pregnancy loss' outcome category. If a woman had a pregnancy loss and a live birth following two or more embryo transfers in the first complete cycle, then we took 'live birth' as the outcome. The HFEA routinely collects information submitted by all licenced UK fertility clinics about their patients, treatments and outcomes.

Outcome

The primary outcome measure for this study was cumulative probability of a live birth for a couple starting their second complete cycle of IVF. A complete cycle included a fresh embryo transfer and associated frozen embryo transfers resulting from a single episode of ovarian stimulation.

Statistical analysis

We developed and temporarily validated a clinical prediction model to use before the start of a second complete cycle of treatment.

For a couple about to embark on their second complete cycle of IVF or ICSI, we estimated the cumulative probability of live birth over this second cycle plus the next two complete cycles using the couple characteristics at the start of their second complete cycle, number of eggs retrieved at the first complete cycle and outcome of the first complete cycle.

Descriptive analysis

Descriptive statistics of the baseline characteristics of the couples and their treatment before initiating their second complete cycle of IVF were presented. Continuous variables with a symmetric distribution were presented as mean (SD), whereas continuous variables with a skewed distribution were presented as median (quartiles). All categorical variables were reported as frequency (percentage).

Model development

Initially, the linearity was examined between the following continuous predictors and the log odds of live birth: woman's age, duration of infertility, time interval between the first and the second egg retrievals, treatment year and the number of eggs retrieved at the first complete cycle. As non-linear associations were found between these continuous variables (except duration of infertility) and live birth, these were then modelled using restricted cubic splines. To aid interpretation, effect estimates of these predictors were presented in the model for the interquartile range (25th versus 75th centile values) (Steyerberg, 2019).

A multivariable discrete time logistic regression model was used to predict the cumulative chance of a live birth from the second to the fourth complete cycle of IVF. The complete cycle number was treated as the time variable for predicting the chance of a live birth in a specific complete cycle and was conditional on no live birth having occurred in the previous cycle.

Based on knowledge from the existing literature, complete cycle number, woman's age and duration of infertility were included in the final model and were not subjected to a variable selection process (Van Loendersloot *et al.*, 2010). For the remaining predictors, a manual backward variable selection process was conducted where Akaike's Information Criteria was used to decide the best fitting model.

To estimate the cumulative predictions of live birth we used the Inverse Probability Weighting (IPW) approach. This approach assumes that women who drop out of treatment after each complete cycle may have a different chance of a live birth compared to women who continue. It attempts to recreate the data as if all couples had completed four complete cycles (Modest *et al.*, 2018). A detailed description of the application process of IPW approach in the model is presented in Supplementary Data S1.

Missing data

The predictor 'duration of infertility (years)' had some missing values. The characteristics of couples with missing data for duration of infertility were compared to those of couples with complete data for this predictor. A Multiple Imputation by Chained Equations (MICE) technique was performed by generating a single imputed dataset to minimize biases and enhance statistical efficiency arising from excluding women with missing predictor values (Greenland and Finkle, 1995). This technique assumed that any missing data were missing at random which means that the missing data depends on the values of the observed predictors and treatment outcome. We had no reason to believe that this missing data would be structurally different from the observed data.

Internal validation

To assess overfitting of the model, an internal validation was carried out using a bootstrap resampling technique. It calculated the optimism-adjusted c-statistic (discrimination) and optimism-adjusted calibration slope. For a more detailed description of this method, please see Supplementary Data S1.

External validation

A prediction model may not perform as well in new patients as in the development set (Steyerberg, 2019). To test the accuracy of the predictions from the model, a temporal validation was performed in a recent UK IVF cohort.

The predictor 'duration of infertility (years)' had missing values in the external dataset and multiple imputation was performed to impute values for this predictor with missing information (Sterne *et al.*, 2009). In this process, 10 imputed datasets were created. This technique assumed that missingness was random conditional on the observed data.

Predictive performance The predictive performance of the model was evaluated in terms of discrimination and calibration. Discrimination assesses the ability of the model to accurately distinguish between women who achieved a live birth and those who did not (Moons *et al.*, 2012). To measure model discrimination, the area under the receiver operating characteristic curve, also known as the c-statistic, was calculated which ranges from 0.5 to 1. A c-statistic of 1 implies perfect discrimination, whereas a c-statistic of 0.5 indicates that the model does not discriminate at all (Hanley and McNeil, 1982). In this study, the c-statistic (and 95% CI) was calculated using the methods suggested by Harrell *et al.* (1996).

To assess the degree of agreement between the observed live birth and predicted probability of a live birth from the model, calibration

was used (van Calster et al., 2019). Calibration can be evaluated by several techniques, such as, calibration-in-the-large, calibration slope and calibration plot. In accordance with the original formula, for the calibration plot, we used the weighted observed CLBR against the weighted predicted probability of live birth and the weights were re-calculated by a newly fitted model in the validation cohort. For perfect calibration, the calibration slope and calibration intercept should be 1 and 0, respectively and the calibration plot shows a diagonal line. Please see Supplementary Data S1 for a detailed description of all calibration techniques used in the study.

The c-statistic, calibration-in-the-large and calibration slope of the model were calculated separately for each of the 10 imputed datasets and separate results were pooled using Rubin's rule (Rubin, 2004). However, for the calibration plot, only the first imputed dataset was used as representative of the other nine datasets.

Recalibrating the model As the model demonstrated poor calibration in the validation cohort, to improve performance, the model was recalibrated (Karp et al., 2004; Toll et al., 2008). This was performed on the first imputed dataset (as represents to all other datasets) using the logistic recalibration method (adjustment of the intercept and the regression coefficients using the calibration intercept and calibration slope respectively). Please see Supplementary Data S1 for a detailed description of method.

All statistical analyses were conducted using the software STATA version 16.

Patient involvement

In setting the research question or the outcome measures, or even in developing plans for design or implementation of the study, no patients were involved. No patients were asked to advise on interpretation or writing up of results of the study. We have plan to disseminate the results of this research study to patients affected by fertility issues via national fertility charities and the HFEA.

Results

After completing the first cycle of IVF/ICSI, 49 314 (43%) of the 113 870 women (73 053 of the 184 269 cycles) commenced a second complete cycle of treatment between January 1999 and September 2008 (see Supplementary Fig. S1).

Baseline characteristics at the start of the second IVF/ICSI cycle

The overall individual and treatment characteristics of the 49 314 couples recorded at the beginning of their second complete cycle are presented in Table I. The mean (SD) age of women who started a second complete cycle was 35 years (4.4). The median (quartiles) duration of infertility for the cohort was 5 years (3, 7). The median time interval between the first and second egg retrievals was 6 months (quartiles 4, 12). The majority of couples was diagnosed with male factor infertility (44.9%). The median number of eggs collected at the start of the first ovarian stimulation was 8 (quartiles 4, 12). Of those couples who commenced the second complete cycle of IVF, ~80% did not get pregnant in their first complete cycle, 12% experienced pregnancy loss and 8% had a live birth resulting from their initial complete cycle of IVF treatment.

In the development cohort, data on duration of infertility were missing for 10.6% women. The distribution of baseline characteristics (at the start of the second complete cycle) between individuals with complete and incomplete data are presented in Supplementary Table SI.

Predicting the cumulative chances of live birth

Unadjusted associations between the probability of live birth and woman's age, duration of infertility, treatment year, time interval of egg retrieval between the first and second complete cycles and the number of eggs retrieved at the first complete cycle are presented in Fig. 1. For women starting their second cycle, the chances of having a live birth declined after age 32. A negative linear association was found between the probability of live birth and duration of infertility. The probability of live birth increased with increasing time since first egg retrieval and increasing number of eggs collected at the first complete cycle in a non-linear manner. The probability of live birth also increased for those starting treatment from 2004.

The predictive factors of the cumulative chance of a live birth from the second up to the fourth complete cycles were woman's age, duration of infertility, year of starting the second complete cycle of IVF, type of treatment, tubal infertility, number of eggs retrieved and outcome of the first complete cycle (Table II, Supplementary Table SII, and Supplementary Data S2). Complete cycle number, woman's age, duration of infertility and tubal infertility were all shown to be negatively associated with live birth. Compared to a second complete cycle of IVF, the odds of live birth were 17% lower after a third and 31% lower after a fourth complete cycle (Table II). The odds of a live birth were 46% lower in women aged 38 years compared to those aged 32 years (odds ratio (OR): 0.54, 95% CI: 0.51 to 0.57). Increasing duration of infertility (7 versus 3 years, OR: 0.89, 95% CI: 0.87 to 0.92) and a diagnosis of tubal infertility (OR: 0.90, 95% CI: 0.86 to 0.94) also reduced the odds of a live birth.

Predictors shown to be positively associated with live birth were a positive outcome in the first complete cycle, year of starting the second complete cycle and the number of eggs retrieved in the first complete cycle. A live birth in the first complete cycle doubles the odds of a live birth (OR: 1.97, 95% CI: 1.82 to 2.14), while a pregnancy loss in a first complete cycle was associated with 35% higher odds of a live birth compared to no pregnancy (OR: 1.35, 95% CI: 1.28 to 1.43). A higher number of eggs retrieved at the first complete cycle also increased the odds of treatment success (12 versus 4 eggs, OR: 1.41, 95% CI 1.36 to 1.45). The 'time interval of egg retrieval between the first and second complete cycle' was not shown to be statistically significant predictor of live birth.

Type of infertility (primary versus secondary) was considered as a potential predictor of live birth. However, it was not included in the final model as the variable did not have any significant association with live birth when it was included in the manual backward selection process.

Assessing predictive ability of the model

In the development cohort, the optimism-adjusted c-statistic and calibration slope of the model were 0.66 (95% CI: 0.65 to 0.67) and

Table I Baseline characteristics of couples' and treatment at the start of their second complete cycle of IVF or ICSI in the development and validation cohorts.

Characteristics (Couple), n (%) unless otherwise stated	Development cohort (HFEA 1999–2008)	Validation cohort (HFEA 2010–2016)
Number of couples, n	49 314	39 442
Number of complete cycles, n	73 053	55 673
Woman's age (year), mean (SD)	35.0 (4.4)	35.6 (4.1)
Duration of infertility (year), median (quartiles)	5 (3–7)	9 (7–11)
Missing	5206 (10.6)	37 306 (95)[a]
Time interval between the first and second egg retrievals (months), median (quartiles)	6 (4–12)	9 (5–24)
Causes of infertility		
Tubal	11 298 (22.9)	6112 (15.5)
Anovulatory	6456 (13.1)	4487 (11.4)
Male factor	22 163 (44.9)	16 318 (41.4)
Unexplained	13 727 (27.8)	11 707 (29.7)
Endometriosis	3215 (6.5)	2939 (7.5)
More than one	7565 (15.3)	4780 (12.1)
Year of second complete cycle started		
1999–2001	9661 (19.6)	NA
2002–2004	16 364 (33.2)	NA
2005–2008	23 289 (47.2)	NA
Type of treatment		
IVF	24 739 (50.2)	15 684 (39.8)
ICSI	24 575 (49.8)	23 757 (60.2)
Number of eggs retrieved at the first complete cycle, median (quartiles)	8 (4–12)	9 (5–13)
Outcome of first complete cycle		
Live birth	3931 (8.0)	4896 (12.4)
Pregnancy loss	5970 (12.1)	6181 (15.7)
No pregnancy	39 413 (79.9)	28 365 (71.9)

[a]Eighty-three percent of women had missing duration of infertility in 2010, increasing to almost 100% in 2017 and in total, 95% of women had missing duration of infertility. HFEA, Human Fertilisation and Embryology Authority.

0.997 (95% CI: 0.97 to 1.04) respectively suggesting moderate discrimination and good calibration. The calibration slope was almost one suggesting no overfitting of the predictor effects of the model.

External validation

For temporal external validation of the model, data were collected from 39 442 women who underwent 55 673 treatment cycles in UK IVF clinics from 2010 to 2016 (see Supplementary Fig. S2). Baseline characteristics of couples and their treatment at the beginning of second complete cycle for the HFEA 2010–2016 cohort are summarized in Table I.

Women included in the validation cohort were 1 year older on average than women in the development sample and had a longer average duration of infertility (Table I). The two predictors 'outcome of first complete cycle' and 'causes of infertility' showed similar distributions across both cohorts. More women underwent ICSI in the validation cohort (60% versus 50%).

Approximately 83% of women in 2010 had missing data on duration of infertility, with this missingness increasing to almost 100% in 2017. Since this variable was mostly complete from 1999 to 2007, data from this time period were used to impute missing values. Multiple imputation of missing data was performed to increase the power of the study by allowing us to include women who would have been excluded otherwise. Before imputation, it has been checked that the continuous predictor 'duration of infertility' was approximately normally distributed.

Assessing predictive ability of the model in an external population

In the validation cohort, the pooled c-statistic, calibration-in-the-large and calibration slope of the original model were 0.65 (95% CI: 0.64 to 0.65), −0.08 and 0.85 (95% CI: 0.81 to 0.88) respectively (see Supplementary Table SIII). The calibration-in-the-large demonstrates that the model systematically over predicted live birth and the average overestimation of prediction of the model was 0.08. And the calibration slope suggests that the original regression coefficient estimates of the model were too large, resulting in extreme predictions for new

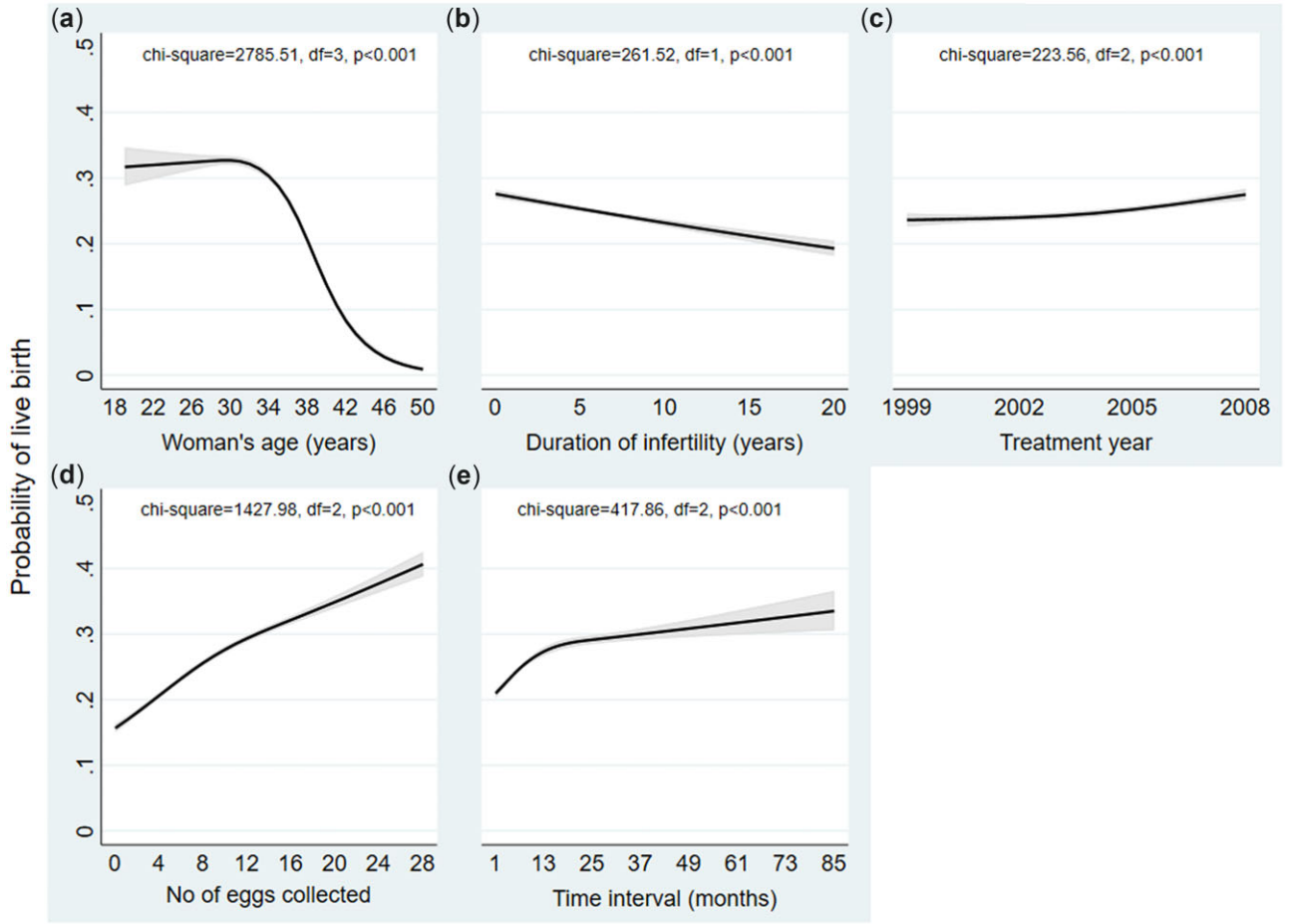

Figure 1. Graphs showing unadjusted (univariable) association between the following continuous baseline variables and a live birth in the second complete cycle of IVF. (**a**) woman's age (years), (**b**) duration of infertility (years), (**c**) year of second complete cycle started, (**d**) number of eggs retrieved at the first complete cycle and (**e**) time interval (months) between egg retrievals in the first and second complete cycles. Each panel depicts the probability of live birth (solid curve) with 95% CIs as a function of the baseline variable.

patients (an over-optimism of around 15%, e.g. low chances of live birth calculated by the model are too low and high probabilities are too high compared with the observed live birth rates).

Figure 2a shows the calibration plot performed on the first imputed dataset (representative of all 10 imputations) depicting the observed cumulative live birth in the validation cohort versus the cumulative predicted probability of live birth from the model applied to the validation cohort (see Supplementary Table SIII). The calibration plot of the model has a calibration intercept of −0.22 (95% CI: −0.26 to −0.19) and a calibration slope of 0.85 (95% CI: 0.81 to 0.88). This plot confirms that the model systematically overestimated the CLBR over three cycles of IVF in the last four deciles of risk. This is seen where the data points are plotted below the reference line. However, the model appears to have estimated probabilities in the rest of the deciles correctly as can be seen by the CIs overlapping the reference line.

To improve calibration, the model was recalibrated by subtracting the calibration intercept of 0.22 and by multiplying the regression coefficients of the original IPW model with the calibration slope of 0.85 (see Supplementary Data S1 and S3; Fig. 2b). The recalibrated regression coefficients are presented in Supplementary Table S2.

Examples of predicting cumulative live birth using the recalibrated model for specific couples

Figure 3 presents an example of cumulative live birth predictions from the recalibrated model for different case scenarios. This shows the cumulative predicted probabilities of live birth over three complete cycles (from the second up to the fourth complete cycle) of IVF for women aged 32 and 40 with either 2 or 5 years of infertility, tubal infertility or not, and 4 or 12 eggs retrieved. The predicted probabilities are presented for each of the three outcome groups of the first complete cycle (live birth, pregnancy loss and no pregnancy).

The cumulative probabilities of live birth are higher in couples who had a live birth in their first complete cycle compared to those who experienced pregnancy loss or had no pregnancy. Again, younger women across all three groups have a much higher chance of success than their older counterparts. The predicted probability of a live birth

Table II Effect of each predictor on live birth over multiple cycles of IVF adjusted for couples' and treatment characteristics at the beginning of second complete cycle based on the Inverse Probability Weighting approach.

Predictors	Odds ratios (95% CI)
Intercept	
Complete cycle number	
2	1
3	0.83 (0.79 to 0.87)
4	0.69 (0.65 to 0.74)
Woman's age*	
38 versus 32 years	0.54 (0.51 to 0.57)
Duration of infertility (years)	
7 versus 3 years	0.89 (0.87 to 0.92)
Time interval of egg retrieval between the first and second complete cycle (months)*	
11 versus 4 months	0.96 (0.93 to 1.00)
Outcome of first complete cycle	
Live birth versus no pregnancy	1.97 (1.82 to 2.14)
Pregnancy loss versus no pregnancy	1.35 (1.28 to 1.43)
Tubal infertility	
Yes versus no	0.90 (0.86 to 0.94)
Year of second complete cycle started*	
2006 versus 2002	1.18 (1.15 to 1.22)
Type of treatment	
ICSI versus IVF	1.05 (1.01 to 1.09)
Number of eggs retrieved at the first complete cycle*	
12 versus 4	1.41 (1.36 to 1.45)

*To aid interpretation of continuous predictors such as woman's age, duration of infertility, time interval of egg retrieval between the first and second complete cycle and year of second complete cycle started, interquartile odds ratio was calculated. It is defined as the ratio of the odds of a live birth for the 75th centile and the odds of a live birth for the 25th centile of the predictor.

for a woman with the most favourable characteristics: 32 years old, 2 years non-tubal infertility, 12 eggs collected at the first ovarian stimulation, and a live born baby in the first complete cycle, is 46% in the second cycle, cumulatively increasing to 81% over three further complete cycles of IVF. In contrast, for women in the same group (live birth) with poorer prognosis, e.g. a woman aged 40 with 5 years tubal infertility and with only 4 eggs collected at her first ovarian stimulation—the predicted probabilities are 18% and 41%, respectively using the recalibrated model.

Couples with the same characteristics who either had a pregnancy loss or no pregnancy at all in their first complete cycle of IVF have lower cumulative predictions than those who had a live birth. For example, for a woman with a good prognosis (age 32, 2 years infertility, no tubal infertility and 12 eggs retrieved) who experienced pregnancy loss in the first complete cycle, the probability of a live birth after the second complete cycle is 38%, increasing cumulatively to 72% over four complete cycles. For a woman sharing the same characteristics but who did not get pregnant at the first complete cycle, these chances are 32% and 64%, respectively. The lowest chances of live birth are predicted for a woman who is aged 40 years, with 5 years tubal infertility, had 4 eggs retrieved at the first complete cycle, and did not get pregnant during the first complete cycle. Her predicted probability of live birth is 11% after the second and 26% over four complete cycles using the recalibrated model. See Supplementary Data S2 and S3 for the formula used to calculate the individualized cumulative predictions.

Discussion

Main findings

Our IVF prediction model can estimate the individualized cumulative chance of live birth over multiple complete cycles for a woman who is about to commence her second complete cycle of IVF or ICSI. Our findings show that woman's age, the outcome of the first complete cycle (live birth, pregnancy loss and no pregnancy), and the number of eggs collected at the first complete cycle are the most significant predictors of subsequent live birth. For example, higher cumulative probabilities of live birth were noted in women who had a live birth in their first complete cycle, compared to those who suffered a pregnancy loss or who had no pregnancy. The model was externally validated using a recent extract of the HFEA data. After updating, the model had good calibration and moderate discrimination in the validation dataset, meaning it could be used in clinical practice in the UK now. This study was conducted and reported in line with TRIPOD guidelines to ensure transparency and quality (Collins *et al.*, 2015).

Strengths and weaknesses of the study

The main strength of this study is its novelty and practical implications for clinical practice. To date, this is the first model that can predict the cumulative chances of live birth in all couples who have finished one full cycle of IVF or ICSI and are about to begin their second complete cycle. A recent USA-based prediction model is able to predict live birth in women starting a second complete cycle but only in women whose first complete cycle was unsuccessful, and it has not been validated on UK data (McLernon *et al.*, 2022). However, our model is applicable to all women starting a second complete cycle and takes account of the outcome of the first complete cycle. This addresses a key unmet need in terms of informing couples of their likely chances of future IVF success after an initial failed cycle of IVF or ICSI in terms of either a pregnancy loss or not becoming pregnant at all. Further, it addresses an unmet need with respect to informing couples who have had a successful first complete cycle of IVF and who wish to further expand their family (Malchau *et al.*, 2017; Paul *et al.*, 2020).

A further strength of this study is that the prediction model considered an IPW method for handling the optimistic assumption of time-to-event models that women who discontinue treatment at different cycles still have the same chance of live birth as those who continue treatment. By this method, the model was weighted to account for a different probability of live birth among those who discontinue treatment. As a result, the model provides more realistic prediction estimates of the individualized chance of a live birth over multiple complete cycles of treatment.

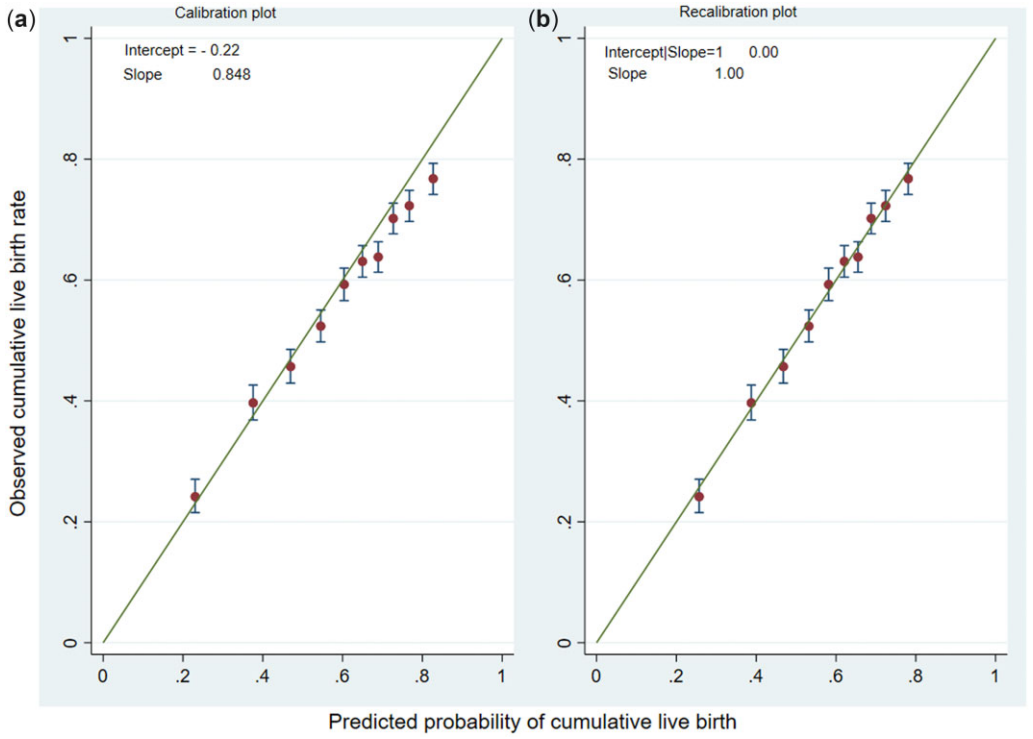

Figure 2. Calibration plots of the association between the predicted and observed cumulative live birth over three IVF/ICSI cycles. Plots illustrate calibration following application of the: (**a**) original model to the validation cohort and (**b**) recalibrated model (after readjusting the intercept and slope) to the validation cohort.

Upon external validation, the reported c-statistic for the model [0.65 (95% CI: 0.64 to 0.65)] suggested moderate discrimination. In general, the c-index for prediction models in reproductive medicine is rather low due to the homogeneity of the study population (e.g., infertile women of reproductive age) (Coppus *et al.*, 2009). To improve calibration, the model was updated. The recalibrated model offers couples individualized predictions of live birth over multiple complete cycles of IVF before embarking on their second complete cycle.

This study has some limitations. For developing the model, 14- to 23-year-old data were used and for validating about 6 to 12 years old data were used. Although the model was updated using this latter dataset, IVF practice continues to evolve over time, which may affect the accuracy of the model. Due to the limited availability of covariate information in the HFEA dataset, the model was not adjusted for some potentially important predictors. These factors include BMI of the women, paternal age, smoking and alcohol intake of the couples, ethnicity and measures of ovarian reserve, such as antral follicle count (Joesbury *et al.*, 1998; Soares and Melo, 2008; Rossi *et al.*, 2011; Marsidi *et al.*, 2021). However, as these measures are often self-reported, there is a possibility that their inclusion may not have necessarily resulted in more reliable estimates (Boniface *et al.*, 2014).

For model development, multiple imputation was performed with one dataset. This is because internal validation using the bootstrap resampling technique, which is the recommended approach to assess model performance, is still an under-researched area in the presence of multiply imputed datasets (Bartlett and Hughes, 2020). However, as the dataset used for developing the models was very large and the amount of missing data was relatively low, the risk of underestimation of the uncertainty associated with imputed values was assumed to be low.

Duration of infertility had a high proportion of missing data in the validation cohort due to a policy of not collecting this information by the HFEA from October 2007. In order to increase statistical power and adjust for any biases caused by excluding women with missing information, multiple imputation was performed assuming that the missingness was random and conditional on the observed data (Sterne *et al.*, 2009; Enders, 2010; White *et al.*, 2011; Van Ginkel, 2020). However, since the variable was consistently recorded between 1998 and 2007, we were able to impute the missing values using all patient data from this time period.

Interpretation of findings in context of the existing literature

The model developed in the study expands on the current literature and extends the use of clinical prediction models in IVF or ICSI treatment. Unlike current prediction models, this novel model is able to predict the cumulative chances of live birth over multiple complete cycles of IVF in couples who are about to start their second complete cycle of IVF or ICSI (McLernon *et al.*, 2016b).

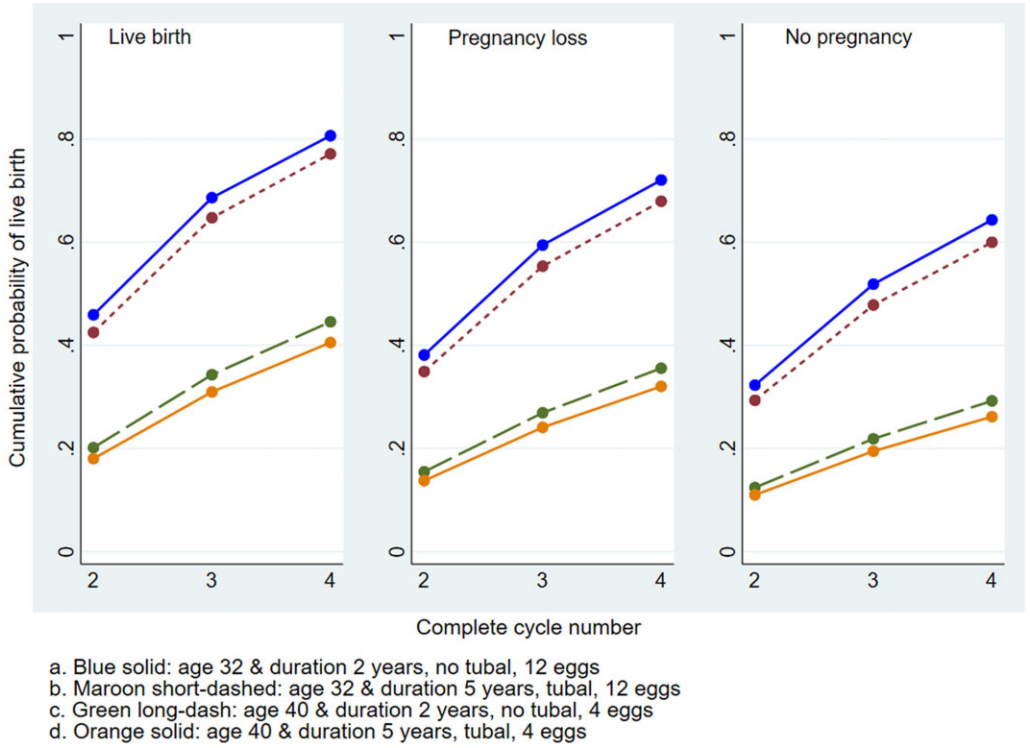

Figure 3. Examples of the recalibrated model predicting CLBR from the second to the fourth complete cycle of IVF by three outcome groups of first complete cycle. CLBR, cumulative live birth rate.

This new model developed extends the suite of models developed by McLernon *et al.*, 2016b which provided predictions for couples about to start their first complete cycle of IVF. It will be added to the online OPIS calculator in due course (https://w3.abdn.ac.uk/clsm/opis/) and will be called 'OPIS-2'. The model uses further information, such as outcome of the first complete cycle and the number of eggs collected at the first complete cycle, to further improve predictions. The inclusion of the variable 'outcome of the first complete cycle' is especially important as it is considered one of the strongest predictors of the model. The model identified that along with woman's age and duration of infertility, treatment outcomes of the first complete cycle and the number of eggs collected at the first complete cycle are significantly positively associated with the chance of having a live birth after a second complete cycle. Tubal infertility and the time interval between the first and second complete cycle were also shown to be negatively associated with the chance of a live birth. These findings agree with those of previous studies (Reichman *et al.*, 2013; Cameron *et al.*, 2017; Paul, 2020). A possible cause is the prejudicial effect of tubal pathology—especially hydrosalpinges—on IVF success and the beneficial effect of salpingectomy (Melo *et al.*, 2020).

The effect estimates of these key prognostic factors of live birth are realistic. This is because the model was weighted accounting for women who drop-out of treatment assuming that the drop-out women could have had a different chance of live birth as compared to women who continued treatment. Predictions from the model that do not account for women who drop out of treatment are too optimistic. Almost all existing previous models were developed considering the conventional optimistic assumptions which resulted in over optimistic predictions (Malizia *et al.*, 2009).

Clinical importance

Since IVF or ICSI treatment success is generally well below 50% per complete cycle, most couples undertake multiple complete cycles, with many discontinuing treatment before having a baby (Gameiro *et al.*, 2012). This is the first model to predict individualized chances of live birth from the second up to the fourth complete cycle of IVF, accounting for couples who discontinue treatment at different treatment cycles. These predictions are relevant not only for individual couples and their clinicians but also for funders and policy makers in determining access to state or insurance funded IVF (National Institute for Health and Care Excellence, 2018).

According to the National Institute for Health and Care Excellence (NICE) guidelines, women in the UK who are under 40 years old should be offered three complete cycles of IVF treatment by the National Health Service (NHS) if: (i) they have been trying to conceive through regular unprotected intercourse for 2 years, or (ii) they have not been able to conceive after 12 cycles of artificial insemination. During the treatment, if a woman reaches the age of 40 the current cycle should be completed but no further cycles should be offered (National Institute for Health and Care Excellence, 2018). In reality, the local Clinical Commissioning Groups (CCGs) across the UK make

their own independent decisions regarding access to IVF funding. In some parts of England (e.g. in the North of England) women are eligible for up to three funded complete cycles of IVF. In other parts of England, most CCGs offer only one funded complete cycle, and some do not provide any funding at all (Human Fertilisation and Embryology Authority, 2018). From 2017, the Scottish Government began funding three complete cycles of IVF for eligible women (National Infertility Group Report, 2016). Therefore, for many couples who do not have access to IVF funding after one complete cycle an estimation of their predicted chances of having a baby if they were to continue treatment would be helpful. Clinicians can use this predictive tool to counsel couples who are about to embark on a second complete cycle of IVF. The tool will inform the couple of their probability of live birth following a second complete cycle. In addition, the tool can provide an estimate of the cumulative probability of live birth after three and four complete IVF cycles which will be useful to inform discussion with couples who wish to plan ahead. Overall, the tool will enable couples to prepare emotionally, financially, and logistically for IVF treatment (Habbema *et al.*, 2004).

It should be noted that the data used in this study contain information on all fertility treatments across the UK and therefore, the findings are particularly relevant for use in the UK. It should also be noted that the predictions of the model are based on historical observational data from couples, not RCTs and therefore, should not be used for decisions regarding treatment effectiveness.

Further research

Since IVF or ICSI treatment practice and policy vary across countries, the geographical external validation of the model using national registry data from other countries would be a useful future research project. Moreover, the model provides more accurate predictions by accounting for women who discontinue treatment at different cycles. Reasons for discontinuing treatment are likely to be varied. They may include women discontinuing treatment due to: 'poor prognosis' (i.e. women who do not have the biological ability to conceive), lack of funds, separation from their partner or having a live birth at the previous cycle. Understanding and collecting information on the reasons for discontinuation of treatment and loss to follow-up is important and should be included in future data collections to improve predictions.

Conclusion

We have developed a novel clinical prediction model that can estimate individualized cumulative chances of live birth in couples starting a second complete cycle of IVF. Woman's age, duration of infertility, treatment outcome of the first complete cycle and number of eggs collected in the first complete cycle were shown to be important predictors of subsequent live birth. Use of this model will address an unmet need in terms of informing couples of their likely chances of future IVF success after completion of an initial IVF cycle.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data underlying this article cannot be shared publicly due to ethical/privacy reasons. The data can be shared on reasonable request to the corresponding author with permission of Human Fertilisation and Embryology Authority.

Authors' roles

D.J.M. and S.B. generated the initial research idea and designed the study. M.B.R. conducted the statistical analysis and literature search and wrote the first draft of the article. D.J.M. supervised the statistical analysis. M.B.R., S.B., N.v.G. and D.J.M. contributed intellectually to the writing and revising of the manuscript. All authors approved the final version of the study.

Funding

This work was supported by the Elphinstone scholarship scheme at the University of Aberdeen and the Assisted Reproduction Unit at Aberdeen Fertility Centre, University of Aberdeen. The funder did not have any role in the study design; data collection, data analysis and interpretation of data; the writing of the report; nor the decision to submit the paper for publication.

Conflict of interest

The authors have no conflict of interest.

References

Abuzeid MI, Bolonduro O, La Chance J, Abozaid T, Urich M, Ullah K, Ali T, Ashraf M, Khan I. Cumulative live birth rate and assisted reproduction: impact of female age and transfer day. *Facts Views Vis Obgyn* 2014;**6**:145–149.

Bartlett JW, Hughes RA. Bootstrap inference for multiple imputation under uncongeniality and misspecification. *Stat Methods Med Res* 2020;**29**:3533–3546.

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**:1–9.

Cameron NJ, Bhattacharya S, Bhattacharya S, McLernon DJ. Cumulative live birth rates following miscarriage in an initial complete cycle of IVF: a retrospective cohort study of 112 549 women. *Hum Reprod* 2017;**32**:2287–2297.

Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed* 2004;**75**:45–49.

Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg* 2015;**102**:148–158.

Coppus SF, van der Veen F, Opmeer BC, Mol BW, Bossuyt PM. Evaluating prediction models in reproductive medicine. *Hum Reprod* 2009;**24**:1774–1778.

Cox DR. Two further applications of a model for binary regression. *Biometrika* 1958;**45**:562–565.

Custers IM, Steures P, Van Der Steeg JW, Van Dessel TJ, Bernardus RE, Bourdrez P, Koks CA, Riedijk WJ, Burggraaff JM, Van Der Veen F *et al*. External validation of a prediction model for an ongoing pregnancy after intrauterine insemination. *Fertil Steril* 2007;**88**: 425–431.

Daya S. Life table (survival) analysis to generate cumulative pregnancy rates in assisted reproduction: are we overestimating our success rates? *Hum Reprod* 2005;**20**:1135–1143.

Enders CK. *Applied Missing Data Analysis*. Guilford Press, 2010.

Gameiro S, Boivin J, Peronace L, Verhaak CM. Why do patients discontinue fertility treatment? A systematic review of reasons and predictors of discontinuation in fertility treatment. *Hum Reprod Update* 2012;**18**:652–669.

Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol* 1995;**142**:1255–1264.

Habbema JDF, Collins J, Leridon H, Evers JLH, Lunenfeld B, te Velde ER. Towards less confusing terminology in reproductive medicine: a proposal. *Hum Reprod* 2004;**19**:1497–1501.

Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**: 29–36.

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**:361–387.

Hernán MA, Hernández-Díaz S, Robins JA. structural approach to selection bias. *Epidemiology* 2004;**15**:615–625.

Hogan JW, Scharfstein DO. Estimating causal effects from multiple cycle data in studies of in vitro fertilization. *Stat Methods Med Res* 2006;**15**:195–209.

Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ Jr. Selection bias due to loss to follow up in cohort studies. *Epidemiology (Cambridge, Mass.)* 2016;**27**:91.

Human Fertilisation and Embryology Authority. 2018. https://www.hfea.gov.uk/treatments/ (13 January 2021, date last accessed).

Joesbury KA, Edirisinghe WR, Phillips MR, Yovich JL. Evidence that male smoking affects the likelihood of a pregnancy following IVF treatment: application of the modified cumulative embryo score. *Hum Reprod* 1998;**13**:1506–1513.

Karp I, Abrahamowicz M, Bartlett G, Pilote L. Updated risk factor values and the ability of the multivariable risk score to predict coronary heart disease. *Am J Epidemiol* 2004;**160**:707–716.

Luke B, Brown MB, Wantman E, Stern JE, Baker VL, Widra E, Coddington IC, 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**:744–752.

Maheshwari A, McLernon D, Bhattacharya S. Cumulative live birth rate: time for a consensus? *Hum Reprod* 2015;**30**:2703–2707.

Malchau SS, Henningsen AA, Loft A, Rasmussen S, Forman J, Nyboe Andersen A, Pinborg A. The long-term prognosis for live birth in couples initiating fertility treatments. *Hum Reprod* 2017;**32**: 1439–1449.

Malizia BA, Dodge LE, Penzias AS, Hacker MR. The cumulative probability of liveborn multiples after in vitro fertilization: a cohort study of more than 10,000 women. *Fertil Steril* 2013;**99**:393–399.

Malizia BA, Hacker MR, Penzias AS. Cumulative live-birth rates after in vitro fertilization. *N Engl J Med* 2009;**360**:236–243.

Marsidi AM, Kipling LM, Kawwass JF, Mehta A. Influence of paternal age on assisted reproductive technology cycles and perinatal outcomes. *Fertil Steril* 2021;**116**:380–387.

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* 2016a;**31**:572–581.

McLernon DJ, Raja EA, Toner JP, Baker VL, Doody KJ, Seifer DB, Sparks AE, Wantman E, Lin PC, Bhattacharya S *et al*. Predicting personalized cumulative live birth following in vitro fertilization. *Fertil Steril* 2022;**117**:326–338.

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* 2016b;i5735.

Melo P, Georgiou EX, Johnson N, van Voorst SF, Strandell A, Mol BW, Becker C, Granne IE. Surgical treatment for tubal disease in women due to undergo in vitro fertilisation. *Cochrane Database Syst Rev* 2020;**10**:CD002125.

Modest AM, Wise LA, Fox MP, Weuve J, Penzias AS, Hacker MR. IVF success corrected for drop-out: use of inverse probability weighting. *Hum Reprod* 2018;**33**:2295–2301.

Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;**98**: 691–698.

National Infertility Group Report. 2016. https://www.gov.scot/publications/national-infertility-group-report/ (13 January 2021, date last accessed).

National Institute for Health and Care Excellence. Fertility Problems: Assessment and Treatment NICE Guidelines [CG156] 2013. Updated 2018. [Online] https://www.nice.org.uk/guidance/cg156 (9 May 2021, date last accessed).

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* 2020;**35**: 1432–1440.

Ratna MB, Bhattacharya S, Abdulrahim B, McLernon DJ. A systematic review of the quality of clinical prediction models in in vitro fertilisation. *Hum Reprod* 2020;**35**:100–116.

Reichman DE, Chung P, Meyer L, Greenwood E, Davis O, Rosenwaks Z. Consecutive gonadotropin-releasing hormone-antagonist in vitro fertilization cycles: does the elapsed time interval between successive treatments affect outcomes? *Fertil Steril* 2013; **99**:1277–1282.

Rossi BV, Berry KF, Hornstein MD, Cramer DW, Ehrlich S, Missmer SA. Effect of alcohol consumption on in vitro fertilization. *Obstet Gynecol* 2011;**117**:136–142.

Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, 2004.

Soares SR, Melo MA. Cigarette smoking and reproductive function. *Curr Opin Obstet Gynecol* 2008;**20**:281–291.

Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;**338**:b2393.

Steyerberg EW. *Clinical Prediction Models*. Cham: Springer International Publishing, 2019.

Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol* 2008;**61**:1085–1094.

Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW; On behalf of Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019;**17**:1–7.

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**:825–831.

Van Der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Hompes PG, Broekmans FJ, Van Dessel HJ, Bossuyt PM, Van Der Veen F, Mol BW; CECERM study group (Collaborative Effort for Clinical Evaluation in Reproductive Medicine). Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. *Hum Reprod* 2007;**22**:536–542.

Van Ginkel JR, Linting M, Rippe RC, van der Voort A. Rebutting existing misconceptions about multiple imputation as a method for handling missing data. *J Pers Assess* 2020;**102**:297–308.

Van Loendersloot LL, Van Wely M, Limpens J, Bossuyt PM, Repping S, Van Der Veen F. Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis. *Hum Reprod Update* 2010;**16**:577–589.

Vrtacnik U, Vrtacnik Bokal E, Devjak R. Cumulative delivery rate after providing full reimbursement in vitro fertilization programme: a 6-years survey. *Biomed Res Int* 2014;**2014**:1–8.

White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011;**30**:377–399.

Wiegerinck MA, Bongers MY, Mol BW, Heineman MJ. How concordant are the estimated rates of natural conception and in-vitro fertilization/embryo transfer success? *Hum Reprod* 1999;**14**:689–693.

Xie J, Liu C. Adjusted Kaplan–Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med* 2005;**24**:3089–3110.