Predicting the chances of having a baby with or without treatment at different time points in couples with unexplained subfertility

Running title: Dynamic prediction of live birth in subfertile couples

D. J. McLernon[1*], A. J. Lee[1], A. Maheshwari[2], R. van Eekelen[3,4], N. van Geloven[5], H. Putter[5], M. J. Eijkemans[4], J. W. van der Steeg[6], F. van der Veen[3], E. W. Steyerberg[5,7], B. W. Mol[8], S. Bhattacharya[9]

[1]Medical Statistics Team, Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, AB25 2ZD, UK; [3]Aberdeen Centre for Reproductive Medicine, Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, AB25 2ZL, UK; [3]Centre for Reproductive Medicine, Academic Medical Centre, 1105 AZ Amsterdam, the Netherlands; [4]Department of Biostatistics and Research Support, University Medical Centre Utrecht - Julius Centre, 3508 GA Utrecht, the Netherlands; [5]Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, 2300 RC Leiden, the Netherlands; [6]Department for Obstetrics and Gynaecology, Jeroen Bosch Ziekenhuis, 's-Hertogenbosch, the Netherlands; [7]Department of Public Health, Erasmus MC - University Medical Centre Rotterdam, 3015 CN Rotterdam, the Netherlands; [8]The Robinson Institute - School of Medicine, University of Adelaide, Adelaide, Australia; [9]Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, AB25 2ZD, UK.

*Corresponding author: DJ McLernon

Address: Medical Statistics Team, Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen, UK, AB25 2ZD.

Email: d.mclernon@abdn.ac.uk

Tel: 01224 437152

Abstract

Study question: Can we develop a prediction model which can estimate the chances of conception leading to live birth with and without treatment at different points in time in couples with unexplained subfertility?

Summary answer: Yes a dynamic model was developed which predicted the probability of conceiving under expectant management and following active treatments (IVF, intrauterine insemination with ovarian stimulation (IUI+SO), clomiphene) at different points in time since diagnosis.

What is known already: Couples with no identified cause for their subfertility continue to have a realistic chance of conceiving naturally which makes it difficult for clinicians to decide when to intervene. Previous fertility prediction models have attempted to address this by separately estimating either the chances of natural conception or the chances of conception following certain treatments. These models only make predictions at a single point in time and are therefore inadequate for informing continued decision making at subsequent consultations.

Study design, size, duration: A population-based study of 1,316 couples with unexplained subfertility attending a regional clinic between 1998 and 2011.

Participants, setting, methods: A dynamic prediction model was developed which estimates the chances of conception within six months from the point when a diagnosis of unexplained subfertility was made. These predictions were recomputed each month to provide a dynamic assessment of the individualised chances of conception whilst taking account of treatment status in each month. Conception must have led to live birth and treatments included clomiphene, IUI+SO, and IVF. Predictions for natural conception were externally validated using a prospective cohort from The Netherlands.

Main results and the role of chance: 554 (42%) couples started fertility treatment within two years of their first fertility consultation. The natural conception leading to live birth rate was 0.24 natural conceptions per couple per year. Active treatment had a higher chance of conception compared to those who remained under expectant management. This association ranged from weak with clomiphene to strong with IVF [clomiphene, Hazard Ratio (HR)=1.42 (95% CI 1.05 to 1.91); IUI+SO, HR=2.90 (2.06 to 4.08); IVF, HR=5.09 (4.04 to 6.40)]. Female age and duration of subfertility were significant predictors, without clear interaction with the relative effect of treatment.

Limitations, reasons for caution: We were unable to adjust for other potentially important predictors, e.g. measures of ovarian reserve, which were not available in the linked Grampian dataset which may have made predictions more specific. This study was conducted using single centre data meaning that it may not be generalizable to other centres. However, the model performed as well as previous models in reproductive medicine when externally validated using the Dutch cohort.

Wider implications of the findings: For the first time, it is possible to estimate the chances of conception following expectant management and different fertility treatments over time in couples with unexplained subfertility. This information will help inform couples and their clinicians of their likely chances of success which may help manage expectations, not only at diagnostic workup completion but throughout their fertility journey.

Study funding/competing interest(s): This work was supported by a Chief Scientist Office postdoctoral training fellowship in health services research and health of the public research (ref PDF/12/06). BWM is supported by a NHMRC Practitioner Fellowship (GNT1082548). BWM reports consultancy for ObsEva, Merck and Guerbet. None of the other authors declare any conflicts of interest.

Key words: Clinical prediction models, live birth, unexplained infertility, subfertility, in vitro fertilisation

76
77
78

Introduction

Infertility or subfertility, defined as the failure to achieve a clinical pregnancy after at least 12 months of regular unprotected sexual intercourse (Zegers-Hochschild et al., 2009), affects around one in six couples (Oakley et al., 2008; Mascarenhas et al., 2012; Zhou et al., 2017). In approximately a third of them the cause of subfertility remains unexplained after routine investigations. Although many will conceive naturally, couples are often unaware of this and equate their condition to sterility (te Velde and Cohlen, 1999; Gnoth et al., 2003; Hunault et al., 2004; van der Steeg et al., 2007). This has encouraged many to seek early access to assisted reproductive technology (Kamphuis et al., 2014), resulting in a steady increase in the number of IVF cycles each year (Ferraretti et al., 2017; Fitzgerald et al., 2017). For example, in the UK, the proportion of all IVF cycles undertaken for unexplained subfertility has increased from 17.5% in 2000 to 27.6% in 2014 (HFEA, 2008; HFEA, 2016).

A few clinical trials of moderate quality have questioned the strategy of active treatment using either clomiphene citrate, intrauterine insemination with ovarian stimulation (IUI+SO) or IVF in couples with unexplained subfertility with reasonably high rates of spontaneous conception (Bhattacharya et al., 2008; Pandian et al., 2015; Brandes et al., 2011; Tjon-Kon-Fat et al., 2016; Bahadur et al., 2016). However, such couples do not have the same chances of conceiving and a previous prediction model has been used in The Netherlands to identify those with an initial poor prognosis (Hunault et al., 2004). Other prediction models developed so far have attempted to estimate the probability of live birth following certain treatments at a single point in time, but their clinical utility has been limited by their inability to update the chances of conception in the same couples at different time points (Leushuis et al., 2009). As couples with better prognoses become pregnant, those remaining behind represent an increasingly biased group with poorer prognoses. Reapplication of a model, developed at an earlier point when couples have better prognosis, to this group of patients has a tendency to overestimate predicted success (van Eekelen et al., 2017). So

far, only one study has made repeated predictions over time of the chances of natural conception (van Eekelen et al., 2017).

Another reason why the ability to assess the prognosis of a couple with unexplained subfertility at different points in time is critical is because it influences decisions around whether, and when, active treatment should be offered (McLernon et al., 2014; ESHRE Capri Workshop Group, 2017). Previous models were unable to assess both the individualised chances of natural conception and conception related to different fertility treatments together - mainly due to the lack of suitable integrated follow up data. By estimating how a couple's chance of conception evolves over time, both naturally and following different fertility treatments, clinicians will have a better understanding of how to manage such patients.

Our aim was to develop and validate a prediction model which can generate dynamic estimates of the chances of conception leading to live birth in couples with unexplained subfertility using a unique record-linked observational database of fertility diagnosis, fertility treatment and pregnancy outcome records (Pandey et al., 2014). We intended that the model should also take into account any potential impact of active treatment by clomiphene, IUI+SO and IVF.

Methods

Study population

This observational study included all couples residing in the Grampian region of Scotland who registered with the Aberdeen Fertility Centre from 1998 to 2011 with unexplained subfertility, i.e. no medically known cause for their barrier to conception following normal tests for tubal patency, semen analysis and ovulation (see supplementary text file 1; fig. S1). We excluded women aged younger than 18 or older than 50 and women who had fertility treatment involving donor eggs or

sperm (fig. S2). We also excluded women who did not give consent for their IVF data to be used for research purposes.

Databases

The following electronic databases were obtained and record-linked:

Aberdeen Fertility Centre Database (AFCD): The AFCD holds all fertility diagnostic and treatment records electronically on all subfertile couples living in the Grampian region of Scotland (Pandey et al., 2014; Sripada et al., 2010; Thomson et al., 2005). Assessments and tests were performed based on the 2004 National Institute for Health and Care Excellence (NICE) guidelines of management of infertility in the UK which have since been updated and, prior to this, Royal College Guidelines (National Collaborating Centre for Women's and Children's Health, 2013). Details of the clinical protocols used in decision-making at Aberdeen Fertility Clinic are included in the Supplementary material (supplementary text file 1). Data entry in the AFCD is validated and checked by regular case note audits.

Assisted Reproduction Unit Database (ARUD): The ARUD contains information on all IVF including intracytoplasmic sperm injection (ICSI) treatment cycles since 1998. The ARUD was record-linked to the AFCD to identify couples who had fertility treatment. ARUD statistics are validated against annual HFEA reports on treatment cycles and outcomes. Access to the AFCD and ARUD was approved by the Aberdeen Fertility Databases Steering Committee.

Aberdeen Maternity and Neonatal Databank (AMND): Live births were identified through record-linkage with the AMND, which has held obstetric records on all women residing in Aberdeen City District since 1950 (Pandey et al., 2014; Maheshwari et al., 2009; Maheshwari et al., 2010). Completeness of the AMND is checked annually against the number of deliveries recorded in the NHS records office. There are several validity checks incorporated within the database to ensure against invalid data entry. Access to the AMND was granted by the AMND Steering Committee.

Scottish Morbidity Records SMR02 Maternity & Neonatal Linked Dataset (SMR02): A proportion of women residing in the Grampian region but outside the catchment area of Aberdeen Maternity Hospital will have delivered in other maternity hospitals. Therefore the SMR02 dataset was obtained which includes maternity inpatients and day-cases receiving care in the obstetrics specialty for the whole of Scotland. Records were extracted from these datasets for the period between 1992 and 2012. Access to the SMR02 linked dataset was approved by the Privacy Advisory Committee of Information Services Division Scotland.

In Scotland, all individuals registered with a general practitioner have a unique identifier, the Community Health Index (CHI) (Evans and MacDonald, 1999). The Data Management Team, University of Aberdeen, securely hold the CHI files for NHS Grampian. The CHI files contain the CHI number and the patient's name, address and date of birth and enable record-linkage of healthcare datasets. All identifiers within the above datasets were removed and replaced with a study specific unique identifier. The pseudonymised datasets were stored within the Grampian Data Safe Haven (DaSH). This process was carried out according to the Standard Operating Procedures of the Data Management Team, University of Aberdeen. The first author (DJM) then record-linked the databases using the unique identifier.

Baseline characteristics

Baseline couple characteristics were extracted from the AFCD and included female age, duration of subfertility, female smoking status (current versus never/ex), female alcohol status (current versus never/ex) and previous pregnancy status at registration. Furthermore, the AFCD provided details (including dates) of treatment with clomiphene citrate and IUI+SO (where the ovarian stimulation is injections (gonadotrophins) alone or a combination of tablets (clomiphene) and injections depending on the individual case) treatments that occurred during follow-up. Details of IVF treatment were obtained from the ARUD.

Outcome

The primary outcome at any point in time, was the occurrence of conception within six months leading to a live birth (shortened to conception for the remainder of this article). The predicted chance of this outcome was first calculated at the point of completion of diagnostic workup (defined as approximately three months after first fertility clinic appointment) and again at every subsequent month until two years after the first appointment (as described below).

Statistical analysis

Descriptive statistics were used to characterise couples at their first fertility clinic appointment. The number (%) of women treated by treatment type up to a maximum of two years and the median time to first treatment were calculated. The frequency of each pattern of sequential treatments was tabulated. The number (%) of conceptions leading to live birth was calculated and grouped into treatment dependent and treatment independent outcomes. The latter group was further split into those occurring without any previous treatment and those following unsuccessful treatment. Conception leading to live birth was classified as treatment dependent if this outcome was recorded in the treatment records or if the 'date of conception' (found by subtracting the gestational age of the baby from the date of delivery) occurred any time during the four weeks prior to date of treatment.

Dynamic prediction

Dynamic prediction modelling using a landmarking approach was conducted to estimate the chances of conception within six months from completion of diagnostic workup (van Houwelingen and Putter, 2012). These chances were updated from every subsequent month until two years after the first fertility clinic appointment (22 times in total). In the dynamic process, the crudest model would involve fitting 22 Cox regression models, each with a separate time origin for each consecutive month with a follow-up window of six months in each. In each model the treatment status would be

updated for that particular month of prediction. As this window slides forward with each successive month any women who conceived in the previous month were excluded from future models. Figure S3 explains the dynamic prediction process for a hypothetical case. Rather than fit 22 separate models which may lead to unstable predictions, we modelled the dynamic process in one 'supermodel' by adding month of prediction as a predictor (linear and quadratic effect) (van Houwelingen and Putter, 2012). Treatment status was the only predictor that was allowed to vary at the month of prediction and was initially coded as: expectant management (never had any treatment - reference category), clomiphene, IUI+SO as the first treatment, IVF as the first treatment, IUI+SO having had previous IUI+SO or IVF since diagnostic workup, IVF having had previous IUI+SO or IVF since diagnostic workup, and, no treatment having had previous treatment since diagnostic workup (which we call 'break from treatment'). However, we found that initial IVF treatment versus continued IVF treatment and initial IUI+SO versus continued IUI+SO had very similar effects and were therefore combined. 'Break from treatment' represents those months in between treatment cycles when the couples under the treatment pathway do not have any treatment. It is essentially a 'nuisance' category that must be included in the model since it is the only monthly treatment status remaining after we consider months of expectant management and fertility treatment. It should be noted, however, that from such a month we can still make a prediction over the following six months during which time they can undergo further treatment as observed in the data. The known predictors of conception female age, previous pregnancy, and duration of subfertility were included in the model at their values at registration. Female age was found to have a non-linear relationship with the log hazard of conception and was fitted as a two-piece linear term i.e. it was modelled as separate linear effects from 18 to 31 years and from 32 to 50 years. Year of registration was included as a linear predictor to account for changes in practice over the 14 year study period. Other available potential predictors, including female smoking and alcohol status and female BMI were added to the initial model but removed if they were not statistically significant. Since female BMI was missing in 20% of women, multiple imputation was

performed to impute these values. BMI was not statistically significant (p=0.99) and was removed from the model. Further details are in the supplementary text file 2.

Predictive ability

We assessed the apparent discriminatory ability of the dynamic prediction model by calculating the concordance statistic for survival time data at each prediction month (Kremers, 2007). Dynamic calibration plots were used to determine whether the predicted probabilities agreed with the observed rates of conception.

External validation

Data from a prospective cohort of patients recruited from 38 hospitals in the Netherlands was used to externally validate predictions for couples under expectant management. The validation cohort contained 5184 patients who were referred for an infertility work up between 2002 and 2004 (van der Steeg et al., 2007; van Eekelen et al., 2017). We selected couples without any major barriers to conception i.e. those with regular menstrual cycles (cycle length between 23 and 35 days), at least one patent fallopian tube and semen analyses with a total motile sperm count $>1 \times 10^6$. Unlike the Grampian record-linked data which followed up women until conception leading to live birth, the Dutch cohort followed couples until natural conception leading to an ongoing pregnancy, defined as the presence of foetal cardiac activity at transvaginal sonography at a gestational age of at least 12 weeks. Therefore, this outcome was used instead of live birth in the validation. If no pregnancy occurred, then follow-up time was censored at the start of treatment or at the last date of contact during follow-up. This meant that we could only validate our model with a cohort following couples for natural conception. The van Eekelen study accounted for missing data by multiple imputation and we used one randomly chosen imputed dataset for validation (van Eekelen et al., 2017).

The parameter estimates of the final model from the Grampian cohort were applied to the validation cohort to assess its predictive ability. Again, the concordance statistic was used as the

measure of discrimination. Calibration-in-the-large (whether the outcome is systematically over or under estimated) was assessed visually by plotting the observed chance of conception within six months (with 95% CI, using Kaplan-Meier estimate) against the average predicted probability of conception within six months for every month since diagnosis. A calibration slope test was conducted to test for overfitting of the model by applying the final model to the Dutch dataset to calculate the linear predictor value for each woman. This linear predictor was then fitted in a new model on its own and its slope estimated. A value significantly below one signifies overfitting.

Secondary analysis

Although the above model provides individualised absolute predictions of conception at every month since diagnosis, the relative treatment estimates are assumed to be the same at each month. We explored whether the treatment estimates vary for couples with different prognosis by fitting interactions between treatment status and each baseline predictor. However, due to the modest sample size this approach leads to overfitting when included with all predictors. An alternative method is to estimate a baseline prognostic score for each couple which is a variable that represents the linear combination of all the predictors in the model (also called the linear predictor or Xß). This is done by fitting a standard Cox regression model predicting time to conception with female age, duration of subfertility, previous pregnancy status and year of registration as predictors. This prognostic score, summarizing all the baseline predictors, was calculated and included in a separate dynamic prediction model as a main effect and as an interaction term with treatment status. Dynamic prediction plots showing the estimated chances of conception by treatment status were presented for couples with poor, average and good prognosis (each defined as the 10^{th}, 50^{th} (i.e. median), and 90^{th} percentile value of the prognostic score respectively). The 95% pointwise confidence limits were estimated from the model and included in the plots.

Statistical analysis was carried out using SAS (V.9.3) (SAS Institute, Cary, NC) and the dynpred package in R (v3.3.3) (R Foundation for Statistical Computing, Vienna, Austria) (Putter, 2015).

Ethical approval: The study was approved by the North of Scotland Research Ethics Committee (12/NS/0120).

Results

A total of 1507 couples were diagnosed with unexplained subfertility after attending AFC for their first consultation between 1998 and 2011 (see fig. S2). We excluded a further 191 couples who conceived during the three month period between their first appointment and completion of workup. The descriptive characteristics of the remaining 1316 couples are shown in Table I. The only substantial set of missing data was for female BMI (missing in 21%).

Out of the 1316 couples, 554 (42%) started treatment within two years of their first appointment. The median (IQR) time to first fertility treatment was 10.9 (5.8, 16.0) months (Table S1). The first treatment was clomiphene in 263 (47%) women. Of these, 147 only had clomiphene treatment, 91 went from a course of clomiphene to IVF (without IUI) and 25 went from clomiphene to IUI+SO. The first treatment was IVF in 184 (33%) women (with 16 subsequently having IUI+SO) and IUI+SO in 107 (19%) women (55 only had IUI+SO and 52 went on to have IVF) (Table S1). Over two years the most frequent treatment pathway was one fresh cycle of IVF (n=73; 13%) (see Table S2 for other treatment pathways). In the month when a patient started clomiphene treatment, the mean (SD) number of clomiphene treatments observed over the following six months was 3.6 (1.8). For IUI and IVF treatments the mean (SD) was 1.7 (0.8) and 1.4 (0.5) respectively. At one year post-workup, a couple who are under the treatment pathway but who happen to have no treatment in that month have a mean (SD) number of treatments over the following six months equal to 0.43 (0.61). In comparison, those still under expectant management at one year post-workup have a mean (SD) number of treatments over the following six months equal to 0.23 (0.66).

From the point of workup completion to 27 months later the conception leading to live birth rate was 0.24 conceptions per couple per year. From the point of treatment to 27 months after workup, the conception leading to live birth rate was 0.39 conceptions per couple per year.

Dynamic prediction of conception leading to live birth

Twenty four women had missing information on duration of subfertility and previous pregnancy status and were excluded from further analysis. For ease of interpretation of treatment predictor effects, Table II presents the hazard ratios with 95% CIs from the final dynamic model. The dynamic prediction model showed that female age at baseline had no significant effect on the likelihood of conception until the age of 32, at which point, the chances decreased by 11% for each yearly increase in age (Hazard Ratio (95% CI) = 0.89 (0.86 to 0.92)). Increasing duration of subfertility was also associated with a decreased chance of conception. The chances of conception in women with no previous pregnancy were 14% lower than in women who had been pregnant before, but this was not statistically significant. The effects of female smoking and female alcohol consumption were weak and hence not included in the final model.

On average, couples who had IUI+SO were almost three times as likely to conceive within six months as couples who were under expectant management (HR=2.90 (2.06 to 4.08)). For couples who had IVF treatment the chances of conception leading to live birth were five times higher compared with expectant management (HR=5.09 (4.04 to 6.40)). The chances of conception following clomiphene (versus expectant management) were only slightly higher (HR=1.42 (1.05 to 1.91)). In the months between treatments, when a couple do not have any treatment, their chances of conception over the following six months were 50% higher compared to couples under expectant management.

Internal validation

The concordance (95% CI) for the dynamic prediction model was 0.60 (0.57 to 0.64) for predictions from workup completion increasing to 0.71 (0.65 to 0.77) for predictions from 21 months post

workup (Table S3). For all treatment groups, the dynamic predictions lay within the 95% CIs of the observed dynamic probabilities (in months where there were enough treated couples to make a Kaplan-Meier estimate).

External validation

The descriptive characteristics of predictors included in the dynamic model were reasonably similar between the Grampian and Dutch cohorts (Table I). From the point of workup completion to 27 months later the ongoing pregnancy rate following natural conception was 0.32 per couple per year. The discriminatory ability of the model applied to the Dutch cohort showed a concordance statistic of 0.63 (0.60 to 0.65) for predictions made at diagnostic workup completion. These ranged from 0.56 to 0.68 over the prediction months (Fig 1A and Table S3). The dynamic calibration plot (see Fig 1B) showed minor evidence of overestimation for predictions at month of workup and at 2 and 16 months post workup. The ratio of observed chance of conception within six months to predicted probability of conception within six months varied from 0.65 at month 16 to 1.06 at month 6. From month three to ten the ratio ranged from 0.95 to 1.06. The calibration slope showed only minor evidence of overfitting (slope=0.968 (95% CI 0.935 to 1.002); p-value for deviation from unity = 0.06).

Predictions for two hypothetical couples

Figure 2 displays the dynamic predictions for two different hypothetical couples. Couple A have been trying to conceive for one year, have never been previously pregnant and the female partner is 30 years old. At each month following diagnostic workup, the blue line represents the probability of conceiving over the following six months for women who were managed expectantly in that month. At the point of workup completion (month 0 in Figure 2) Couple A's chance of conception over the following six months is 31% (95% CI 25% to 37%). The other solid lines on the plot represent the probability of conception over six months when shifting from expectant management to fertility

treatment in a given month. All treatments resulted in an increased chance of conception over the following six months. For example, six months after diagnosis, if they are still under expectant management, Couple A's chances of conception over the following six months reduce to 20% (95% CI 15% to 24%). Their probability would increase slightly to 27% (95% CI 18% to 35%) if the female partner starts clomiphene at month six (result not shown). However, if she starts IUI+SO at month six her chance is 47% (95% CI 31% to 59%) (green line on Fig 2) whereas if she starts IVF at month six her chance of conception within the next six months is 67% (53% to 77%) (red line on Fig 2). Twelve months after diagnostic workup, their chance of conception over the next six months decrease to 17% (13% to 22%) if the couple continues to try naturally. If the couple start treatment with clomiphene, IUI+SO or IVF 12 months after diagnosis their chances of conceiving would be 24% (15% to 31%), 42% (27% to 54%) and 62% (47% to 72%) respectively. See supplementary text file 3 for a demonstration on how to calculate these predictions using the model formula.

Secondary analysis

The chances of conception by treatment status did not vary for couples with different baseline characteristics (i.e. the interactive effect was not statistically significant, p=0.26). Dynamic prediction plots showing the estimated chances of conception by treatment status are presented in Figure 3 for couples with poor, average and good prognosis. IVF (versus expectant management) was shown to have a relatively constant hazard ratio of around five for all three prognostic levels. However, the effect of IUI+SO is stronger for poor prognosis couples and weaker for good prognosis couples.

Discussion

Main findings

This study provides individualised dynamic predictions of the chances of natural and treatment dependent conception leading to live birth for couples with unexplained subfertility. These predictions are estimated from the point of fertility workup completion and are updated monthly thereafter. At each month we were able to predict the impact of having different treatment (clomiphene, IUI+SO and IVF). We found that increasing age of the woman (from 32 years), increasing duration of subfertility and previous pregnancy measured at baseline were predictive of conception leading to live birth. Treatment with either IUI+SO or IVF was associated with a higher chance of conception leading to live birth in comparison with expectant management.

Strengths

This project is unique in that the study cohort came from a single NHS fertility clinic with complete population coverage in the Grampian region of North East Scotland. The clinic has an NHS maternity unit with no provision of either service in the private sector. We successfully record-linked electronic patient registration, diagnostic and treatment data from Aberdeen Fertility Clinic with pregnancy outcome information from maternity hospital admissions in one defined geographical region (Hansen et al., 2015). This provided a unique resource to follow-up couples from registration at the fertility clinic to delivery of a baby (Pandey et al., 2014), and, for the first time, allows us to estimate the chances of conception following different fertility treatments at different time points based on a couple's individual characteristics.

With the exception of a recently published model predicting natural conception only (van Eekelen et al., 2017), previous models have only been able to predict fertility outcomes at a single point in time. While this provides an initial prognosis, it is insufficient for making clinically meaningful decisions about when to initiate fertility treatment in couples who possess a good chance of natural conception. It is far more useful to assess how the prognoses of these couples change over time in order to balance the risks and benefits of a variety of active treatments versus the option of expectant management. The predictions for untreated patients were validated using an external

natural conception cohort of couples with unexplained subfertility from The Netherlands. Whilst this cohort did not contain follow-up information on those who were treated, we were able to validate predictions following expectant management. Whilst the criteria for absence of fertility factors in this validation cohort were not consistent with the criteria used to select cases for the development cohort, the model still performed well which shows that it is robust and transportable to cohorts with slightly different diagnostic processes.

Limitations

First, we were unable to adjust for other potentially important predictors that were not available in the linked Grampian dataset which may have made predictions more specific. These include ethnicity and measures of ovarian reserve such as Anti-Müllerian hormone or follicle stimulating hormone (van Loendersloot et al., 2010). However, we did use the most important predictors of conception, found in previous research, that are known to the clinician before starting treatment. These were used in the Hunault model which estimates the chances of natural conception over a year following diagnosis of unexplained or mild infertility (Hunault et al, 2004). This model is used in the Netherlands to facilitate decisions around who should receive IVF treatment. Its predictors include female age, duration of infertility, previous pregnancy status, sperm motility, and referral status (secondary versus tertiary care). Since we only included couples with unexplained infertility we had no need to adjust for sperm motility. Also, almost all couples in our cohort were referred from primary care (93%) so we did not adjust for this variable. It is possible that we may be missing known and unknown factors that could have confounded the effect of treatment on conception.

Second, since this study was conducted using single centre data, albeit covering a complete region of Scotland, it may not be generalizable to other centres. However, this single clinic serves the entirety of a defined geographical region and the model performed as well as previous models in reproductive medicine when externally validated using the Dutch cohort (van der Steeg et al., 2007). The concordance ranged from 0.57 to 0.71 in the development cohort and from 0.56 to 0.68 in the

validation cohort; previous research has shown that this can be explained by the homogeneous nature of the population and that calibration is a more important measure of performance (Coppus et al., 2009). The fact that we could only source an external cohort with couples under expectant management for validation reflects the uniqueness of the valuable Grampian dataset. The model showed only marginal evidence of overfitting with a calibration slope of 0.97 (instead of 1) and there was slight visible evidence of overprediction in the calibration plot for the first three months.

Third, the modest size of the database and the associated number of treatments commencing in a given month meant that our model may have lacked precision, especially in the first few months for treated couples e.g. 16 couples initiated IUI+SO and 15 initiated IVF. The lower numbers for IUI+SO and IVF are reflected in the wide confidence bands for the predictions in Figure 2 and 3. 18 months after workup, 399 (53.6%) of the remaining 744 women were still under expectant management so precisions for that group remained reasonable over follow-up, which is also reflected in the figures. Our model may have lacked the power to detect whether the relative treatment effect varied by baseline couple characteristics. Due to modest sample size we attempted to explore whether baseline prognostic score (a linear combination of individual baseline predictors) varied the chances of conception following different treatments. This interactive effect was not statistically significant. Future work will involve prospectively following up a suitable cohort of patients to enable us to fully validate the model with sufficient statistical power, including refinements and updating of our model (Steyerberg, 2009). However, the treatment effect does vary at the absolute risk scale which is most important for prediction.

Fourth, following the commencement of fertility treatment, during the six month follow-up period there will be months when couples will not have had further treatment or will have started another treatment. This means that they could achieve conception either as a result of this treatment, naturally, or following any other treatment received during follow-up. The same applies to couples who were under expectant management in a given month but who then received treatment during

the six month follow-up. However, at the time when the prediction is being made (i.e. at the start of each month) we do not have this future information available. Our statistical model reflected these real world pathways observed in the study database.

Fifth, one of the assumptions we make is that couples were trying to conceive naturally during the 'non-treatment' periods and after the last treatment received. For the latter we assume that they still had an active desire to conceive.

Lastly, since we used historical data, the treatment protocols for similar couples will have changed over time, favouring IVF post 2008 (HFEA, 2008; HFEA, 2016). We adjusted for year of registration at the clinic in the model but it was not a strong predictor of conception. However, the predictions from the model that we present in Figure 2 and in supplementary text file 3 are based on the most recent year (i.e. 2012). Therefore predictions will reflect conception chances from couples seen and treatment received in that year. Further, IVF success rates have not improved since 2012 in the UK - in 2012 the live birth rate per treatment cycle was 26% whereas in 2016 it was 25% (HFEA, 2018). Further, other countries still use clomiphene to treat women with unexplained subfertility so these results will be of interest to clinics around the world.

Comparison with previous fertility prediction models

Previous prediction models only provide a static, i.e., once only, prediction of live birth either at the point of diagnosis or at the very start of treatment (Leushuis et al., 2009). The Hunault model (Hunault et al., 2004), for example, estimates the chance of natural pregnancy leading to live birth among subfertile couples within one year from their first attendance at a fertility clinic. It has been validated and is used in the Netherlands in the form of a web-based calculator to facilitate decisions around whether a couple should receive fertility treatment straightaway or undergo expectant management (Kamphuis et al., 2014; Hunault et al., 2005). However, this model cannot be used to estimate how a couple's prognosis changes over time nor to provide predictions following

commencement of different fertility treatments. One cannot simply reapply the Hunault model to the same couple one year later because, by this time, some of the more fertile couples will have conceived in the intervening time leaving a cohort that is less fertile. Reapplying the model would result in over predictions which are insufficiently explained by an increase in female age or duration of subfertility (van Eekelen et al., 2017). This is evidenced in a recent study that estimated the chances of natural conception repeatedly over time (van Eekelen et al., 2017). Conversely, some prediction models only estimate the chances of live birth following IVF (Templeton et al., 1996; Nelson and Lawlor, 2011; van Loendersloot et al., 2013; Luke et al., 2014; Dhillon et al., 2015; McLernon et al., 2016). Couples also need to know what their chance of natural conception is in order to determine whether undergoing such invasive and expensive treatment actually improves their chances. This problem is acknowledged in the NICE fertility guidelines which attempted to assess the cost-effectiveness of IVF treatment using the predicted probabilities from two separate models – one for predicting success following IVF and one for predicting natural conception (Hunault et al., 2004; National Collaborating Centre for Women's and Children's Health, 2013; Nelson and Lawlor, 2011).

Interpretation and clinical importance

Our analyses show that the predicted chance of success after clomiphene was only slightly higher than with expectant management, an outcome which is in agreement with results of previous randomized controlled trials (RCTs) (Hughes et al., 2010). We found that the chance of conception in women who started treatment with IUI+SO was higher than for women remaining under expectant management. A recent Cochrane review showed that there was no evidence of a difference in live birth rate between IUI+SO versus expectant management in couples with unexplained subfertility (Veltman-Verhulst et al., 2016). However, they acknowledge that the quality of the evidence was moderate and no subgroup analysis was presented and our result may be explained partially by IVF cycles occurring following IUI+SO.

Our results indicate that the chance of live birth following IUI+SO or IVF is higher than that associated with expectant management. Does this mean that we should treat everyone immediately (Kamphuis et al., 2014; Kersten et al., 2015)? Clinicians must consider the absolute chances with and without treatment on an individual basis together with the cost to the health service (or patient) and the physical and emotional burden to the woman (Brandes et al., 2011). A 30 year old woman with no previous pregnancy who has been trying to conceive for one year has a 31% chance of conceiving over the next 6 months (and subsequently having a live birth) with expectant management. This 31% chance also reflects the possibility of starting treatment at some point during that six month follow-up period. If we chose a cut-off of ≥20% to indicate a good prognosis over six months then we would wait 6 months when her chances drop below that cut-off before deciding to treat. If she started IVF treatment at month 6 then her chances of conceiving over the subsequent 6 months (which allows for the fact that further treatments may or may not occur during that time) increase to 67%.

Those couples who start treatment will have subsequent months with and without further treatment. When making predictions from a month in which such couples do not have treatment, their chance of conception over the following six months was higher than for couples under expectant management. This is because these couples are more likely to have (further) treatment over the following six months than those under expectant management.

The most appropriate use of our model is to facilitate communication between the clinician and the couple when explaining their chance of conceiving over six months at different points in time when they are under a period of expectant management and how these chances may improve when they commence a particular fertility treatment in the future. Couples can then manage their expectations accordingly. It will also be useful to couples whose first treatment attempt is unsuccessful and who would like to know what their predicted chances are when continuing treatment. It should be noted that these predictions are based on historical observational data from couples, not randomised

controlled trials, and therefore should not be used for decisions regarding treatment effectiveness. The information should be supported by clinical knowledge.

Further research

Whilst the chance of conception was higher with IVF or IUI+SO compared with expectant management, the absolute chances varied from couple to couple depending on their characteristics. Future work should investigate individualised thresholds of predictions to facilitate treatment decisions. This could mean that having a slightly larger chance of conception with treatment may not necessarily lead to treatment when balanced with the associated financial, physical and emotional burden.

We recognise that further validation is necessary in order to strengthen the generalisability and transportability of the model, particularly in relation to predictions following treatment. In the absence of such a database we plan to collect a prospective cohort of couples with unexplained subfertility that will include high quality diagnostic, treatment and outcome information. Following validation, the model may be converted into an online prediction calculator that can be easily used in clinical practice to inform couples of their evolving chances of conception.

The predictions from our model will help clinicians to manage the expectations of couples with unexplained subfertility not only at the diagnostic workup completion but throughout their fertility journey.

Author Contributions: Dr McLernon had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: McLernon, Lee, Bhattacharya, Putter.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: McLernon, Bhattacharya.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: McLernon, Putter.

Obtained funding: McLernon.

Administrative, technical, or material support: McLernon, van Eekelen, van Geloven, Putter, Eijkemans, van der Steeg, van der Veen, Steyerberg, Bhattacharya.

Acknowledgements: We acknowledge the data management support of the Grampian Data Safe Haven (DaSH) and the associated financial support of NHS Research Scotland, through NHS Grampian investment in the Grampian DaSH. For more information, visit the DaSH website http://www.abdn.ac.uk/iahs/facilities/grampian-data-safe-haven.php.

We would like to thank Kar Yee Lor, a medical student at the University of Aberdeen, for searching case notes, and all the staff at Aberdeen Fertility Clinic for their help with database queries and case note searching.

Funding: This work was supported by a Chief Scientist Office postdoctoral training fellowship in health services research and health of the public research (ref PDF/12/06). The views expressed here are those of the authors and not necessarily those of the Chief Scientist Office. The funders had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; or decision to submit the manuscript for publication.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. BWM is supported by a NHMRC Practitioner Fellowship (GNT1082548). BWM reports consultancy for ObsEva, Merck and Guerbet. None of the other authors declare any conflicts of interest.

References

Bahadur G, Homburg R, Muneer A, Racich P, Alangaden T, Al-Habib A, Okolo S. First line fertility treatment strategies regarding IUI and IVF require clinical evidence. *Hum Reprod* 2016;**31**:1141-1146.

Bhattacharya S, Harrild K, Mollison J, Wordsworth S, Tay C, Harrold A. Clomifene citrate or unstimulated intrauterine insemination compared with expectant management for unexplained infertility: pragmatic randomised controlled trial. *Br Med J* 2008;**337**:a716.

Brandes M, Hamilton CJCM, van der Steen JOM, de Bruin JP, Bots RS, Nelen WL, Kremer JA. Unexplained infertility: overall ongoing pregnancy rate and mode of conception. *Hum Reprod* 2011;**26**:360-368.

Coppus SFPJ, van der Veen F, Opmeer BC, Mol BWJ, Bossuyt PMM. Evaluating prediction models in reproductive medicine. *Hum Reprod* 2009;**24**:1774-1778.

Dhillon RK, McLernon DJ, Smith PP, Fishel S, Dowell K, Deeks JJ, Bhattacharya S, Coomarasamy A. Predicting the chance of live birth for women undergoing IVF: a novel pre–treatment counselling tool. *Hum Reprod* 2016;**31**:84-92.

van Eekelen R, van Geloven N, van Wely M, McLernon DJ, Eijkemans MJ, Repping S, Steyerberg EW, Mol BW, Bhattacharya S, van der Veen F. Constructing the crystal ball: how to get reliable prognostic information for the management of subfertile couples. *Hum Reprod* 2017;**32**:2153-2158.

van Eekelen R, Scholten I, Tjon-Kon-Fat, van der Steeg J W, Steures P, Hompes P, van Wely M, van der Veen F, Mol BW, Eijkemans M J, et al. Natural conception: repeated predictions over time. *Hum Reprod* 2017;**32**:346-353.

583 ESHRE Capri Workshop Group. A prognosis-based approach to infertility: understanding the role of
584 time. *Hum Reprod* 2017;**32**:1556–1559.

585 Evans JMM, MacDonald TM. Record-linkage for pharmacovigilance in Scotland. *Br J Clin Pharmacol*
586 1999;**47**:105-110.

587 Ferraretti AP, Nygren K, Nyboe Andersen A, de Mouzon J, Kupka M, Calhaz-Jorge C, Wyns C,
588 Gianaroli L, Goossens V. The European IVF-Monitoring Consortium (EIM), for the European Society of
589 Human Reproduction and Embryology (ESHRE); Trends over 15 years in ART in Europe: an analysis of
590 6 million cycles. *Hum Reprod Open* 2017;**2017**(2): https://doi.org/10.1093/hropen/hox012.

591 Fitzgerald O, Harris K, Paul RC, Chambers GM. Assisted reproductive technology in Australia and New
592 Zealand 2015. Sydney: National Perinatal Epidemiology and Statistics Unit 2017. University of New
593 South Wales Sydney.

594 Gnoth C, Godehardt D, Godehardt E, Frank-Herrmann P, Freundl G. Time to pregnancy: results of the
595 German prospective study and impact on the management of infertility. *Hum Reprod* 2003;**18**:1959-
596 66.

597 Hansen M, de Klerk N, Stewart L, Bower C, Milne E. Linked data research: a valuable tool in the ART
598 field. Hum Reprod 2015; **30**: 2956-2957.

599 HFEA. A long term analysis of the HFEA Register data (1991-2006). 2008.
600 http://www.hfea.gov.uk/docs/Latest_long_term_data_analysis_report_91-06.pdf. Accessed 24 Jan
601 2017.

602 HFEA. Fertility treatment 2014 trends and figures. 2016.
603 http://ifqtesting.blob.core.windows.net/umbraco-website/1783/fertility-treatment-2014-trends-
604 and-figures.pdf. Accessed 24 Jan 2017.

HFEA. Fertility treatment 2014-2016 trends and figures. 2018. https://www.hfea.gov.uk/media/2563/hfea-fertility-trends-and-figures-2017-v2.pdf. Accessed 26 Oct 2018.

van Houwelingen HC and Putter H. Dynamic prediction in clinical survival analysis. 1st edn, 2012. CRC Press, Boca Raton.

Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JLH, te Velde ER. Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod* 2004;**19**:2019-2026.

Hunault CC, Laven JSE, van Rooij IAJ, Eijkemans MJC, te Velde ER, Habbema JDF. Prospective validation of two models predicting pregnancy leading to live birth among untreated subfertile couples. *Hum Reprod* 2005:**20**:1636-1641.

Hughes E, Brown J, Collins JJ, Vanderkerchove P. Clomiphene citrate for unexplained subfertility in women. *Cochrane Database of Systematic Reviews* 2010;1:CD000057. doi:10.1002/14651858.CD000057.pub2.

Kamphuis EI, Bhattacharya S, van der Veen F, Mol BWJ, Templeton A, for the Evidence Based IVF Group. Are we overusing IVF? Br Med J 2014; **348**: g252.

Kersten FAM, Hermens RPGM, Braat DDM, Hoek A, Mol BW, Goddijn M, Nelen WL, Improvement Study Group. Overtreatment in couples with unexplained infertility. *Hum Reprod* 2015;**30**:71–80.

Kremers WK. Concordance for Survival Time Data: Fixed and Time-Dependent Covariates and Possible Ties in Predictor and Time. Technical Report Series #80. Mayo Foundation 2007.

Leushuis E, van der Steeg JW, Steures P, Bossuyt PM, Eijkemans MJ, van der Veen F, Mol BW, Hompes PG. Prediction models in reproductive medicine: a critical appraisal. *Hum Reprod Update* 2009;**15**:537–52.

van Loendersloot LL, van Wely M, Limpens J, Bossuyt PMM, Repping S, van der Veen F. Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis. *Hum Reprod Update* 2010;**16**:577-89.

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**:2972–2980.

Luke B, Brown MB, Wantman E, Stern JE, Baker VL, Widra E, Coddington CC 3rd, Gibbons WE, Ball GD. A prediction model for live birth and multiple births within the first three cycles of assisted reproductive technology. *Fertil Steril* 2014;**102**:744–752.

Maheshwari A, Scotland G, Bell J, McTavish A, Hamilton M, Bhattacharya S. The direct health services cost of providing assisted reproduction in overweight or obese women: a retrospective cross-sectional analysis. *Hum Reprod* 2009;**24**:633-639.

Maheshwari A, Scotland G, Bell J, McTavish A, Hamilton M, Bhattacharya S. Direct health services costs of providing assisted reproduction services in older women. *Fertil Steril* 2010;**93**:527-536.

Mascarenhas MN, Flaxman SR, Boerma T, Vanderpoel S, Stevens GA. National, Regional, and Global Trends in Infertility Prevalence Since 1990: A Systematic Analysis of 277 Health Surveys. *PLOS Med* 2012;**9**(12):e1001356.

McLernon DJ, te Velde ER, Steyerberg EW, Mol BWJ, Bhattacharya S. Clinical prediction models to inform individualized decision-making in subfertile couples: a stratified medicine approach. *Hum Reprod* 2014;**29**:1851-1858.

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. *Br Med J* 2016;**355**:i5735. http://dx.doi.org/10.1136/bmj.i5735.

National Collaborating Centre for Women's and Children's Health. Fertility: assessment and treatment for people with fertility problems 2013. Available at http://www.nice.org.uk/nicemedia/live/14078/62769/62769.pdf.

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.

Oakley L, Doyle P, Maconochie N. Lifetime prevalence of infertility and infertility treatment in the UK: results from a population–based survey of reproduction. *Hum Reprod* 2008;**23**:447–50.

Pandey S, McLernon DJ, Scotland G, Mollison J, Wordsworth S, Bhattacharya S. Cost of fertility treatment and live birth outcome in women of different ages and BMI. *Hum Reprod* 2014;**29**:2199-2211.

Pandian Z, Gibreel A, Bhattacharya S. In vitro fertilisation for unexplained subfertility. Cochrane Database of Systematic Reviews 2015;11:CD003357. doi:10.1002/14651858.CD003357.pub4.

Hein Putter. dynpred: Companion Package to "Dynamic Prediction in Clinical Survival Analysis". 2015. R package version 0.1.2. https://CRAN.R-project.org/package=dynpred.

Sripada S, Townend J, Campbell D, Murdoch L, Mathers E, Bhattacharya S. Relationship between semen parameters and spontaneous pregnancy. *Fertil Steril* 2010;**94**:624-630.

van der Steeg JW, Steures P, Eijkemans MJC, Habbema JD, Hompes PG, Broekmans FJ, van Dessel HJ, Bossuyt PM, van der Veen F, Mol BW, et al. Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. *Hum Reprod* 2007;**22**:536-542.

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

673 Templeton A, Morris JK, Parslow W. Factors that affect outcome of in–vitro fertilisation treatment.
674 *Lancet* 1996:**348**:1402–1406.

675 Thomson F, Shanbhag S, Templeton A, Bhattacharya S. Obstetric outcome in women with
676 subfertility. *Br J Obstet Gynecol* 2005;**112**:632-637.

677 Tjon-Kon-Fat RI, Bensdorp AJ, Scholten I, Repping S, van Wely M, Mol BWJ, van der Veen F. IUI and
678 IVF for unexplained subfertility: where did we go wrong? *Hum Reprod* 2016;**31**:2665-2667.

679 te Velde ER, Cohlen BJ. The management of infertility. *New Engl J Med* 1999;**340**:224-225.

680 Veltman-Verhulst SM, Hughes E, Ayeleke RO, Cohlen BJ. Intra-uterine insemination for unexplained
681 subfertility. *Cochrane Database of Systematic Reviews* 2016;2:CD001838.
682 doi:10.1002/14651858.CD001838.pub5.

683 Zegers-Hochschild F, Adamson GD, de Mouzon J, Ishihara O, Mansour R, Nygren K, Sullivan E,
684 Vanderpoel S. The International Committee for Monitoring Assisted Reproductive Technology
685 (ICMART) and the World Health Organization (WHO) Revised Glossary on ART Terminology, 2009.
686 *Hum Reprod* 2009;**24**:2683–2687.

687 Zhou Z, Zheng D, Wu H, Li R, Xu S, Kang Y, Cao Y, Chen X, Zhu Y, Xu S, et al. Epidemiology of infertility
688 in China: a population-based study. *Br J Obstet Gynecol* 2017; doi:10.1111/1471-0528.14966.

689

690

691 **Figure legends**

692 **Figure 1. Dynamic assessment of the model applied to the external validation cohort**

693 A Discrimination using concordance statistic at each month since diagnosis of unexplained
694 subfertility (95% Confidence Interval (CI)); B calibration plot of observed chance of conception within

six months (95% CI) versus average predicted probability of conception within six months at each month since diagnosis of unexplained subfertility.

Figure 2. Dynamic predictions of conception within six months following expectant management and treatment. Predictions are made at every month following workup completion (up to 21 months) for:

Couple A who have been trying to conceive for 12 months, with no previous pregnancy, where the female partner is 30 years old;

Couple B who have been trying to conceive for 24 months, with previous pregnancy, where the female partner is 40 years old.

Note: The x-axis shows the time in months since the end of diagnostic work-up. At each time point, the prediction shown is the probability of conception over the following six months given the treatment status at that time point. The grey vertical reference lines show the predicted probability of conceiving within six months from six months and twelve months post diagnostic work-up completion. The left hand plot for Couple A is described in the results section. The right hand plot (Couple B) shows the dynamic predictions for a couple who have been trying to conceive for two years, with previous pregnancy, and where the female partner is 40 years old. At the point of their diagnostic workup completion, the couple have a 12% (95% CI 9% to 16%) chance of conception over the next 6 months (and subsequently having a live birth) (see expectant management blue line at month 0). Six months after diagnosis, the couple's chances of conception over the next six months reduce to 7% (95% CI 5% to 10%) (see where blue line crossed the dashed vertical reference line). If the couple start IUI at month six their chance is 20% (95% CI 12% to 28%) (green line) whereas if they start IVF at month six their chance of conception is 33% (22% to 42%) (red line).

Figure 3. Dynamic predictions of conception within six months following expectant management and treatment for couples with poor, average and good prognosis. Predictions are made at every month following diagnostic workup completion (up to 21 months)

Poor prognosis couples were taken as the 10th percentile of the prognostic score. The hazard ratios (95% CI) are: 5.50 (2.69 to 11.23), intrauterine insemination with ovarian stimulation (IUI+SO) versus expectant management and 5.89 (4.01 to 8.65), IVF versus expectant management.

Average prognosis couples were taken as the 50th percentile of the prognostic score. The hazard ratios (95% CI) are: 2.87 (2.04 to 4.04), IUI+SO versus expectant management and 5.07 (3.99 to 6.43), IVF versus expectant management.

Good prognosis couples were taken as the 90th percentile of the prognostic score. The hazard ratios (95% CI) are: 2.00 (1.13 to 3.56), IUI+SO versus expectant management and 4.66 (3.28 to 6.62), IVF versus expectant management.

Table I Characteristics of couples with unexplained subfertility at the time of registration at the fertility clinic

Characteristics	Grampian N (%), unless otherwise stated	The Netherlands N (%), unless otherwise stated
Total number of women	1316	5235
Female age, mean (SD)	33.0 (4.9)	32.5 (4.3)
Female age group, years		
<31	418 (31.8)	1873 (35.8)
31 to 35	485 (36.9)	2239 (42.8)
36 to 40	320 (24.3)	941 (18.0)
>40	93 (7.1)	182 (3.5)
No previous pregnancy (female partner)	745 (56.6)	3439 (65.7)
Duration (mths), median (IQR)	24 (18, 36)	19 (15, 28)

SD=standard deviation; IQR=interquartile range.

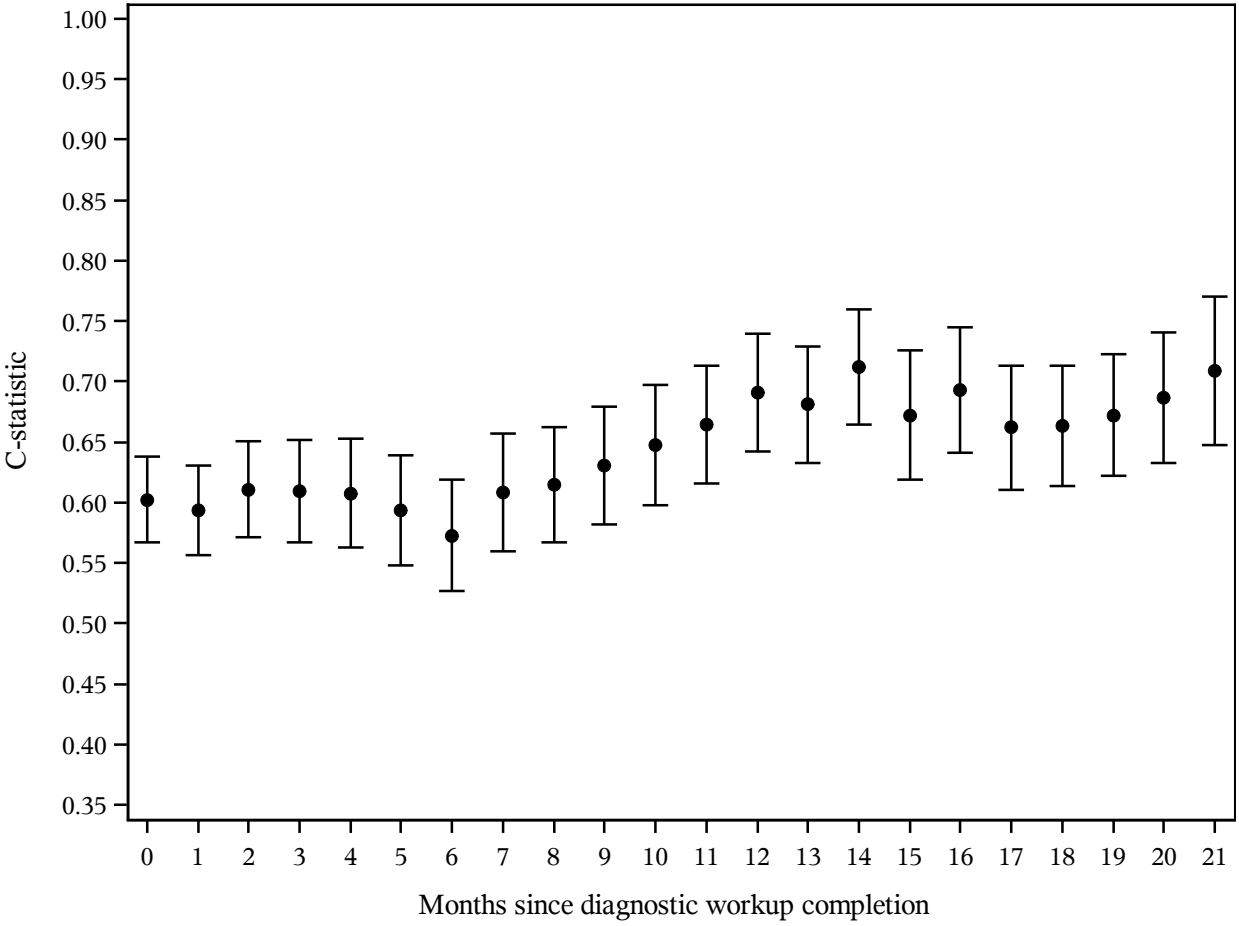
Table II Effect of female and couple characteristics and treatment status during follow-up on conception leading to live birth

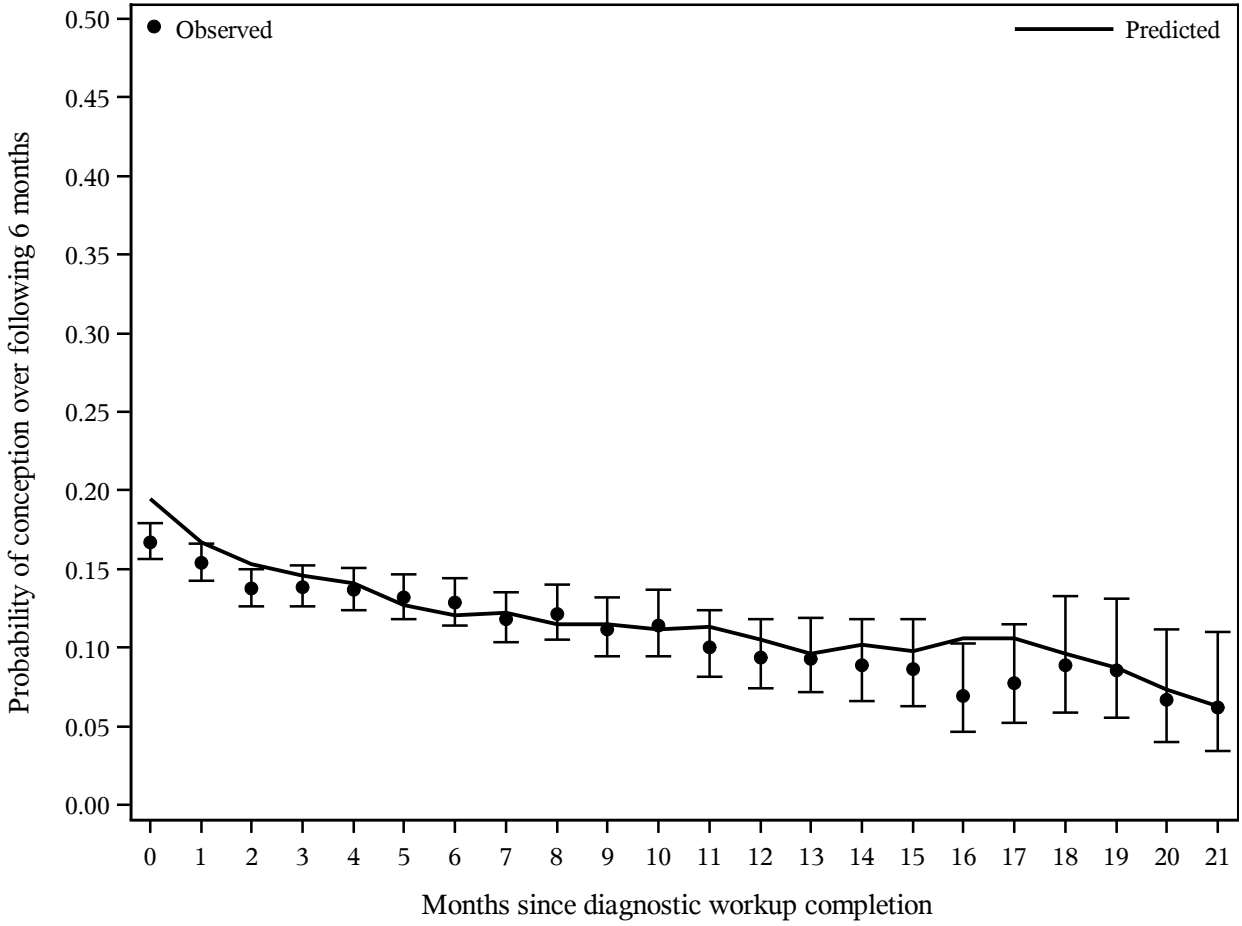
Characteristics	Hazard ratio (95% CI)	p-value
Female age ≤31 years†	1.01 (0.96 to 1.05)	0.83
Female age ≥32 years†	0.89 (0.86 to 0.92)	<0.001
Duration of infertility (years)	0.85 (0.79 to 0.92)	<0.001
Previous pregnancy (no v yes)	0.86 (0.71 to 1.03)	0.11
Year of registration	1.02 (1.00 to 1.05)	0.06
Treatment status in a month		<0.001
Expectant management	Ref	-
Clomiphene	1.42 (1.05 to 1.91)	0.02
IUI+SO	2.90 (2.06 to 4.08)	<0.001
IVF	5.09 (4.04 to 6.40)	<0.001
Break from treatment	1.50 (1.17 to 1.89)	0.001
Prediction month[+]		<0.001
S1[+]	0.00017 (0.00010 to 0.00028)	<0.001
S2[+]	4.02 (2.45 to 6.61)	<0.001

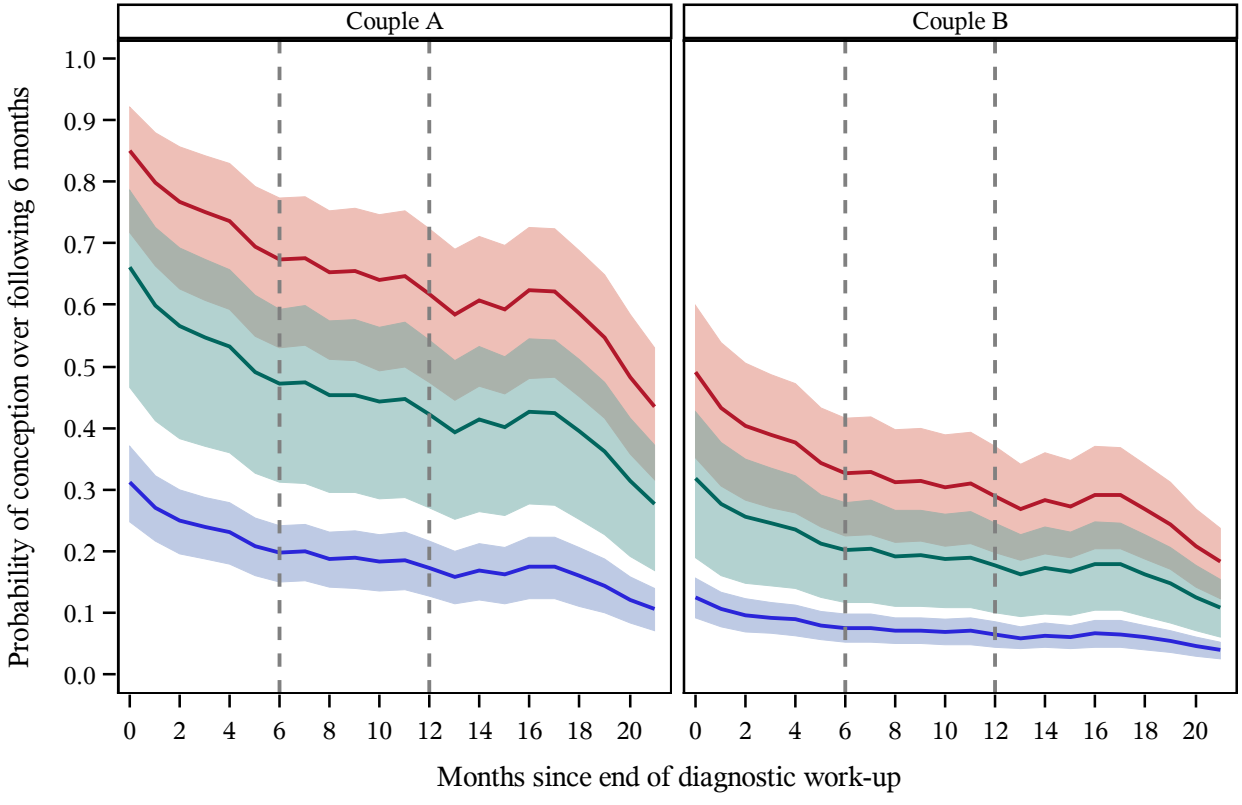
†Female age had a non-linear relationship with the log hazard of live birth and was modelled as two linear predictors.

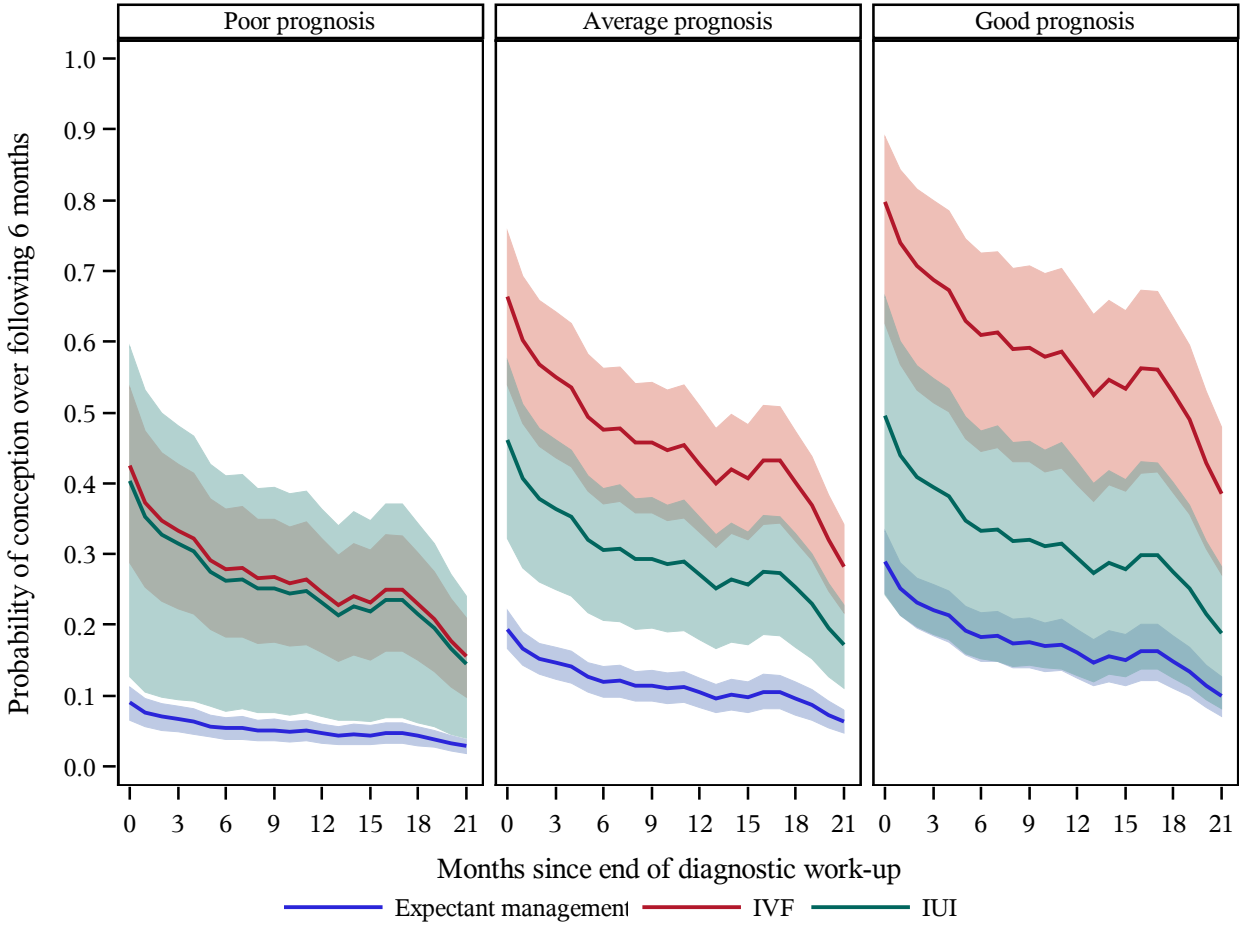
[+]Prediction month represents the effect of the month when dynamic predictions are made, S1 = (Prediction month/640.5); S2 = (Prediction month/640.5)^2

Baseline survival at 6 months = 0.61; IUI+SO = ovarian stimulation intrauterine insemination; IVF = in-vitro fertilisation


C-statistic
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Months since diagnostic workup completion


Observed
Predicted
Probability of conception over following 6 months
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Months since diagnostic workup completion


Couple A
Couple B
Probability of conception over following 6 months
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0 2 4 6 8 10 12 14 16 18 20
Months since end of diagnostic work-up
Expectant management
IVF
IUI
Reference lines for probability of conception within 6 months from 6 and 12 months post work-up


Poor prognosis
Average prognosis
Good prognosis
Probability of conception over following 6 months
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0 3 6 9 12 15 18 21
Months since end of diagnostic work-up
Expectant management
IVF
IUI

Clinical Protocols

Fertility clinic assessment and tests were performed per protocol, as per the evidence based standard operating procedures of the unit based on the NICE guidelines of management of infertility.
Once a couple was referred by the GP to the fertility clinic they were seen by a clinician and a detailed history and examination was performed in the fertility clinic along with a trans vaginal scan to assess the uterus, the ovaries and any adnexal pathology. Patient had a mid-luteal progesterone measurement to confirm ovulation. A single semen analysis was required for all men but and was repeated if the first result was abnormal. If the patient had an irregular menstrual cycle and features suggestive of PCOS further endocrine investigations such as measurement of gonadotropins, prolactin, thyroid and adrenal function, including androgen levels, was done. Rubella immunity was established for all women and Chlamydia screening performed if the woman was less than 30 years of age. Obese women were referred to the dietician for assessment and weight loss advice. Tests for tubal patency were organised in appropriate cases. Laparoscopy and Dye test was the investigation of choice for the assessment of pelvic anatomy including tubal patency. HSG was preferred when a less invasive approach to assess the pelvis was preferred. The couple then came for a follow-up consultation where the test results were discussed and based on the diagnosis further appropriate treatment was organised.

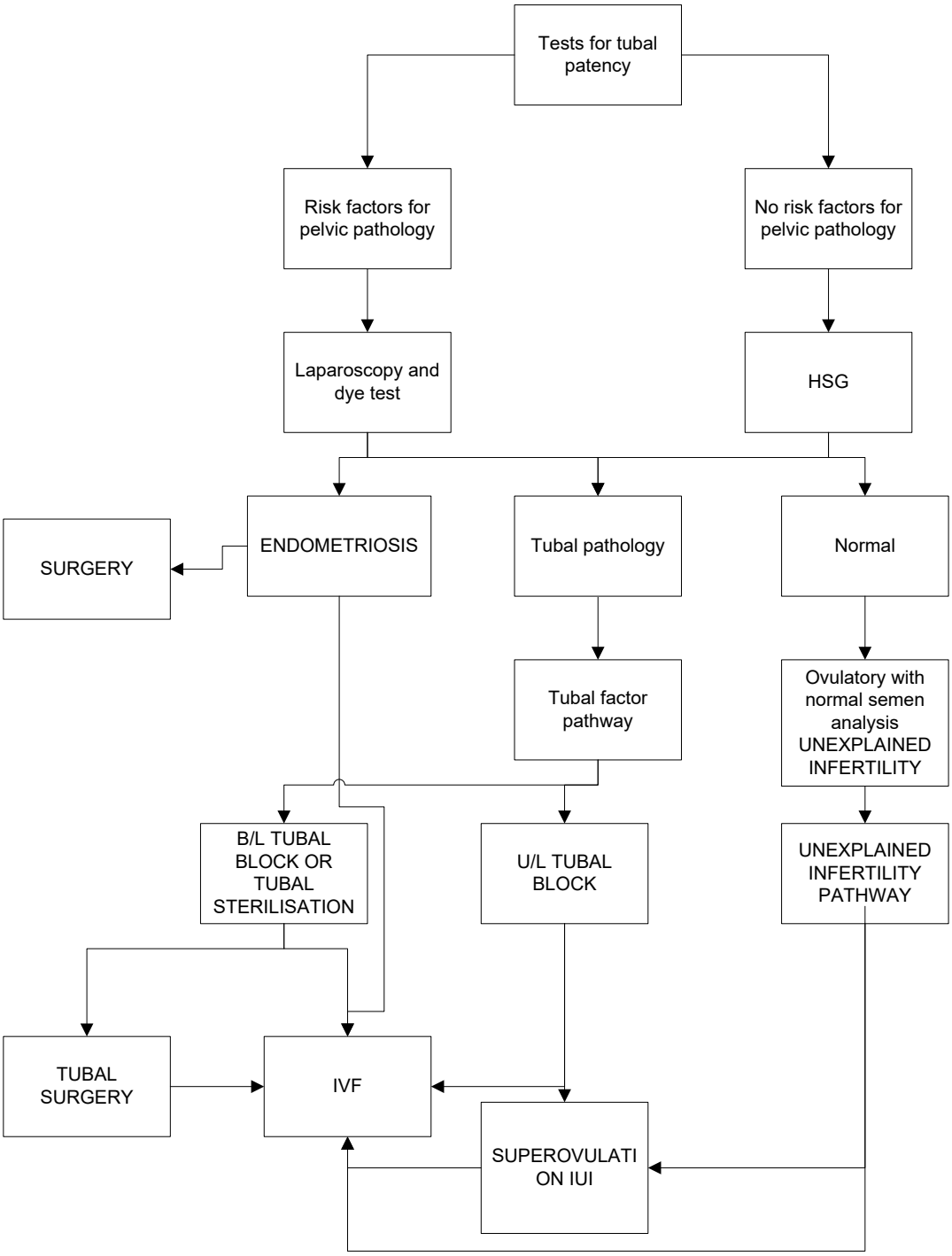
Ovulation induction (Clomiphene/Gonadotropins)
Women with BMI >30 kg/m2 were given a weight loss target of 5-10% and had a consultation with the nurses for Clomiphene chat. Once the weight loss target was achieved if they were still amenorrhoeic and the pregnancy test was negative, Progesterone was given for a withdrawal bleed. Clomiphene 50mg was started from day 2-6 of the cycle USS was performed on day 12 of first cycle and day 21 progesterone was checked to confirm ovulation. In the absence of ovulation after 2 cycles the dose was increased. In the absence of pregnancy after 4 ovulatory cycles a tubal patency assessment was organised and patient was added to IVF waiting list. If tubal patency assessment was normal clomiphene was continued for 12 cycles. After 12 ovulatory cycles 3 cycles of gonadotropin IUI was offered. If the patient remained anovulatory after 4 cycles of clomiphene, gonadotropin IUI was offered for 6 cycles.

Women needing Gonadotrophin induction of ovulation were referred for hour long nurse appointment where the procedure was discussed in detail. Gonadotrophin was commenced after a baseline USS and blood test. Low dose step up regimen was used. Monitoring was done with serial blood tests and USS. When 1-2 follicles were 17mm HCG trigger was administered. IUI was performed 30-36 hrs later.

Super ovulation IUI
Super ovulation IUI was offered to ovulatory women with mild male factor or unexplained infertility. Patient had appointment with doctor and nurse for consultation. After normal baseline scan and estradiol levels FSH was administered alternate days and monitoring was performed by blood tests and scans from day 8-9. HCG trigger was administered when 1-2 follicles were >17mm and IUI was performed 36 hrs later.

IVF
A standard long protocol for ovarian stimulation was used for IVF during this period. The initial stimulatory dose of follicle-stimulating hormone (FSH) was determined by the age of the patient (150 IU in women less than 37years old; 225 IU in 37–39 year olds and 300 IU in women 40 years and above). Women with polycystic ovaries were started on 150 IU of FSH. Monitoring during the period of study was performed by ultrasound as well as serum estradiol. A human chorionic gonadotrophin (HCG) trigger was administered once at least three follicles of more than 17–18 mm diameter were visualized by ultrasound scan. Double-embryo transfer was the routine policy during the period to which the data relates (this has now changed to selective single-embryo transfer).

Fig. S1 Diagnostic and fertility treatment pathway for tubal patency



Methods: Further details of the statistical analysis

In the dynamic modelling process, the crudest statistical model involved fitting 22 Cox regression analyses, each with a separate time origin for each consecutive month with a follow-up window of six months in each. As this window slides forward with each successive month any women who conceived (leading to live birth) in the previous month(s) were excluded from future models. Rather than fit 22 separate models, we modelled the dynamic process in one 'supermodel' by adjusting for a function of month. First of all this involves creating a 'super prediction dataset' using the following steps (van Houwelingen and Putter, 2012):

1. Fix a prediction window, w =six months;
2. Select a set of prediction time points, called landmarks, in this case every month from 3 to 24 months;
3. Create a prediction dataset for each landmark by truncation and administrative censoring;
4. Stack all datasets into a single super prediction dataset.

The variables that change in each of the datasets in point 4 are their respective landmark number, the treatment type occurring in that month (clomifene, IUI, IVF treatment gap versus expectant management), and the censoring indicator. Known predictors of live birth, such as female age, previous pregnancy history, and duration of subfertility were included in the model at their baseline values. Year of registration was included as a predictor to account for changes in practice over the 14 year study period. Other potential predictors such as female smoking status and female alcohol status and female BMI were included in the model but removed if they were not statistically significant. The chance of conception was known to decline from the female age of 32 years from previous studies (Hunault et al, 2004) and this non-linear relationship was confirmed in Martingale residual plots. Therefore female age was modelled as a two-piece linear effect. The super Cox model was adjusted for month of prediction as a predictor and a robust sandwich estimator was used to adjust the standard errors for the fact that one couple contributed information in multiple landmarks, even though only a single conception (leading to live birth) was considered as the endpoint. Since female BMI was missing in 20% of women, multiple imputation was performed to impute these values. This procedure assumes that missing data were missing at random, conditional upon the observed predictors and outcome. BMI was not statistically significant (p=0.99) and was removed from the model meaning that the imputed databases were abandoned and only the original dataset used.

Predictive ability

Discrimination assesses a prediction model's ability to correctly distinguish between patients who achieve a pregnancy (for whom the model assigns a high probability) and do not achieve a pregnancy (for whom the model assigns a low probability). We assessed the apparent discriminatory ability of the dynamic prediction model using a concordance statistic for survival time data (Kremers 2007). The concordance statistic allows for tied events in time and was calculated for each landmark and its six month follow-up time. Dynamic calibration plots were created by plotting the average predicted probability of a pregnancy over six months for each landmark month by treatment type and treatment history status. The plot was then overlaid by the observed probability of pregnancy over six months by landmark month calculated using Kaplan-Meier estimator.

References

van Houwelingen HC and Putter H. Dynamic prediction in clinical survival analysis. 1st edn, 2012. CRC Press, Boca Raton.

Kremers WK. Concordance for Survival Time Data: Fixed and Time-Dependent Covariates and Possible Ties in Predictor and Time. Technical Report Series #80. Mayo Foundation 2007.

Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JLH, te Velde ER. Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. Hum Reprod 2004;19:2019-2026.

Figure S2 Flow chart of exclusion criteria

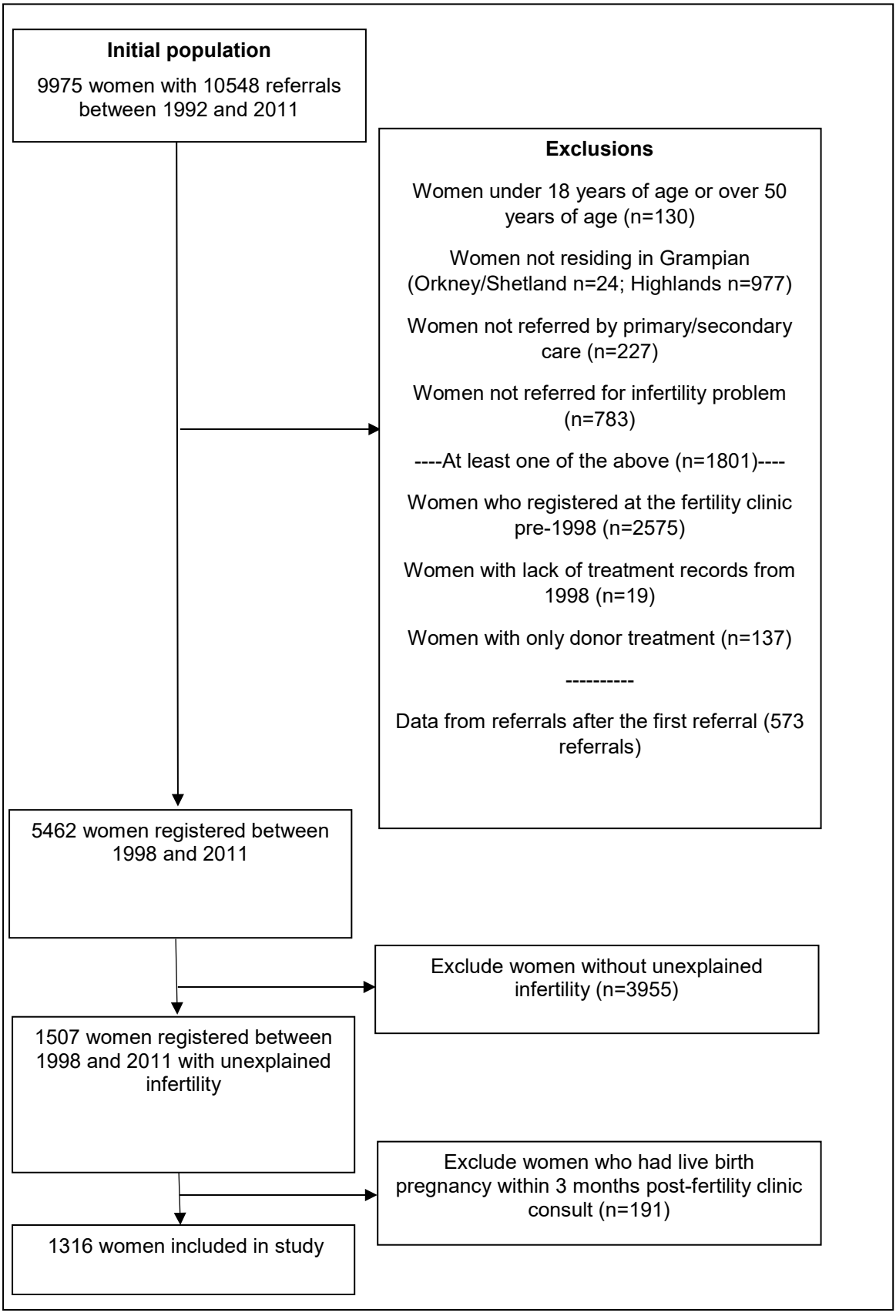
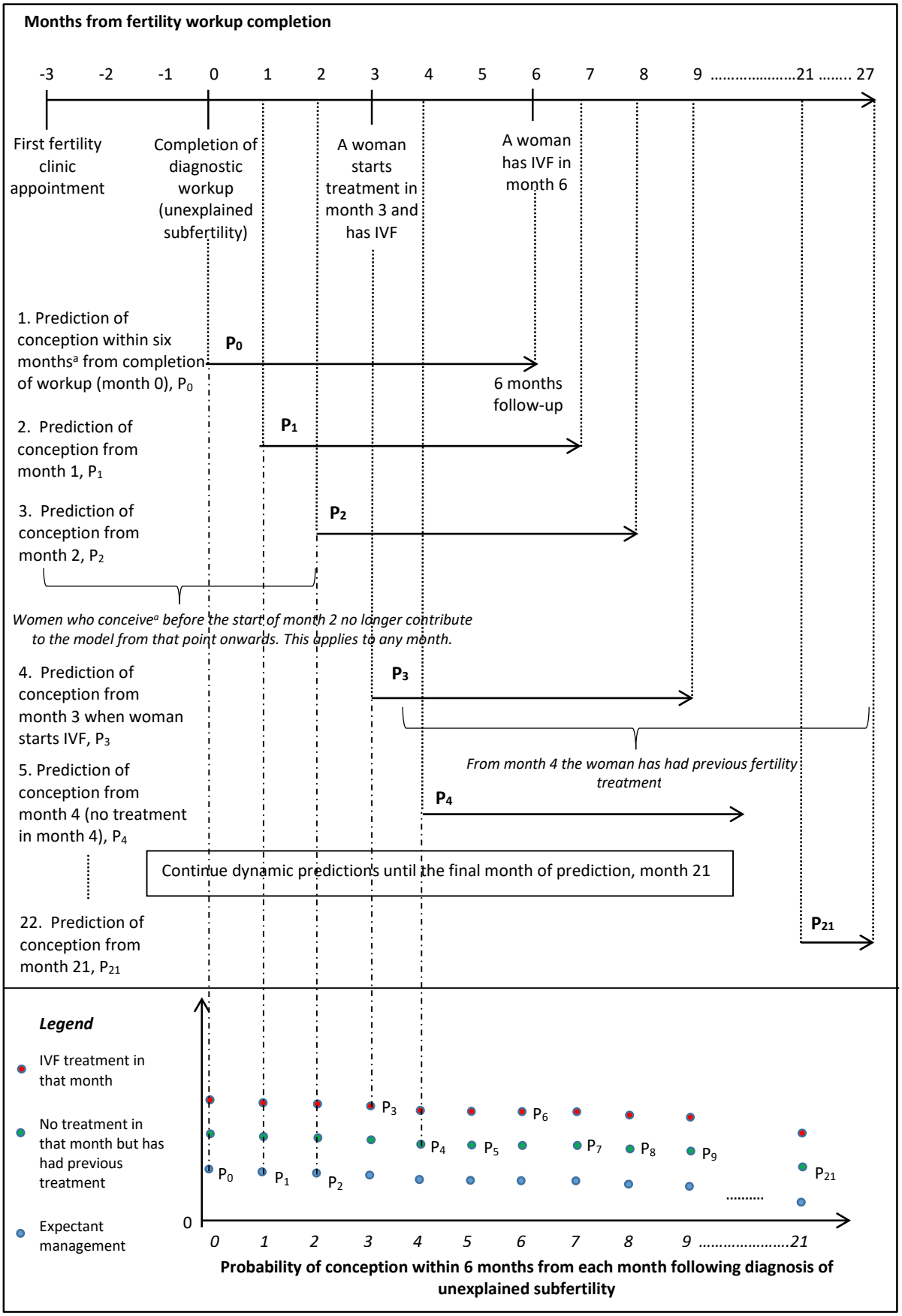


Figure S3 Diagram explaining dynamic prediction of conception using a hypothetical example



Note: The top part of the diagram explains the dynamic prediction of conception within six months from diagnostic workup and every month thereafter. The conception must have ultimately led to a live birth. The lower part of the diagram contains the predicted chances of conception for every possible treatment scenario occurring at every month from completion of diagnostic workup. The above example is for a woman who is under expectant management until three months post workup completion (the blue dots showing the chances of conception over six months from workup completion P_0, one month after workup completion P_1, and from two months after workup completion P_2). At month three she starts IVF treatment (the clinician/patient would then read the predicted chance of conception from the red dot, P_3, at month 3 post workup completion). The woman then has a break from treatment for two months (the clinician/patient would then read the predicted chance of conception from the green dots, P_4 and P_5, at months 4 and 5 post workup completion). She then has her second IVF cycle at month 6 (the red dot at P_6 shows the predicted chance of conception within six months from month 6). Assuming she never has further treatment we would then read all predictions from future months at the green dots from P_7 to P_{22}).

[a] Conception within six months which leads to a live birth.

Table S1 Frequency of treatment and time to first treatment

Treatment	Number who ever had treatment (%)[a]	Number having as treatment within 2 years (%)[a]	Number having as first treatment within 2 years (%)[a]	Time to first treatment in months, median (IQR)
Clomifene	274 (44.4)	265 (47.8)	263 (47.5)	7.1 (4.4 to 11.0)
IUI+SO	159 (25.8)	152 (27.4)	107 (19.3)	11.1 (7.1 to 14.3)
Fresh IVF	398 (64.5)	343 (61.9)	184 (33.2)	13.0 (7.8 to 16.4)
Frozen IVF	94 (15.2)	83 (15.0)	0 (0.0)	-
Any treatment	617 (46.8)	554 (42.1)	554 (42.1)	10.9 (5.8 to 16.0)
No treatment	699 (53.1)	762 (57.9)	762 (57.9)	-

IQR=interquartile range
IUI+SO= intrauterine insemination with ovarian stimulation
[a] For individual treatments the denominator is everyone who had a treatment. For any treatment and no treatment the denominator is all women.

Table S2 Frequency of treatment patterns over the first two years following fertility clinic registration by first treatment

First treatment	Treatment pattern							No. women (%)[a]
Clomifene	Clomifene	Clomifene						30 (11)
	Clomifene	Clomifene	Clomifene					30 (11)
	Clomifene							28 (11)
	Clomifene	Clomifene	Clomifene	Clomifene	Clomifene	Clomifene		20 (8)
	Clomifene	Clomifene	Clomifene	Clomifene	Clomifene			16 (6)
	Clomifene	Clomifene	Clomifene	Clomifene				12 (5)
	Clomifene	Clomifene	Clomifene	Clomifene	Clomifene	Clomifene	Clomifene	7 (3)
	Clomifene	Clomifene	Clomifene	Clomifene	Clomifene	Clomifene	Fresh IVF	6 (2)
	Clomifene	Clomifene	Clomifene	Clomifene	Fresh IVF			5 (2)
	Clomifene	Clomifene	Clomifene	Fresh IVF				5 (2)
	Others							104 (39)
IUI+SO	IUI+SO							25 (23)
	IUI+SO	IUI+SO						18 (17)
	IUI+SO	IUI+SO	IUI+SO					8 (7)
	IUI+SO	IUI+SO	Fresh IVF					8 (7)
	IUI+SO	Fresh IVF						7 (7)
	Others							41 (38)
IVF	Fresh IVF							73 (40)
	Fresh IVF	Fresh IVF						35 (19)
	Fresh IVF	Frozen IVF						14 (8)
	Fresh IVF	Fresh IVF	Fresh IVF					11 (6)
	Fresh IVF	Fresh IVF	Frozen IVF					5 (3)
	Others							46 (25)
Total								554 (100)

IUI+SO= intrauterine insemination with ovarian stimulation

[a] For individual treatment patterns the denominator is everyone who had the corresponding first treatment.

An example of how to calculate dynamic predictions using the model

To make predictions we calculate the following formula: S = 1 - Baseline survival^exp(XB)

where baseline survival is the chance of not conceiving within six months for the reference group (see Table below for the relevant value according to the month of prediction)

Month of prediction	Baseline survival
0	0.6448
1	0.5724
2	0.4670
3	0.3405
4	0.2170
5	0.1357
6	0.06362
7	0.01783
8	0.004304
9	0.0003951
10	0.00002089
11	0.0000001813
12	0.000000001796

B = the log hazard ratios from Table 2; and X = the value of the predictors for each couple.

For example, for Couple A (who have been trying to conceive for 12 months, with no previous pregnancy, where the female partner is 30 years old) we want to calculate their chances of conception at baseline, i.e. workup completion, over the following six months.

1. Using the log hazard ratio information (parameter estimates) for each predictor:

   XB = 0.0051*Female age (≤31) - 0.11*Female age (≥32) - 0.16*duration of infertility (years) - 0.15*previous pregnancy + 0.024*year of registration + (0.35*clomifene + 1.06*IUI+SO + 1.63*IVF + 0.40*break from treatment) - 8.68*S1 + 1.39*S2

   where

Female age (≤31) = actual female age in years if age is less than 32; otherwise if female age is greater than 31 this equals 31;

Female age (≥32) = if female age is less than 32 this equals zero, if female age greater than 31 this equals age – 31;

Duration of infertility = length of time spent trying to conceive prior to first fertility clinic appointment in years;

Previous pregnancy = 1, if never previously pregnant or =0, if previously pregnant;

Year of registration always equals zero so ignore;

Treatment (a maximum of only one of the following four treatment groups must equal 1 at any month of prediction):

If having clomifene in month of prediction then set clomifene=1, otherwise 0

If having IUI+SO in month of prediction then set IUI+SO =1, otherwise 0

If having IVF in month of prediction then set IVF =1, otherwise 0

If having break from treatment in month of prediction then set break from treatment =1, otherwise 0

If under expectant management in month of prediction then set all four of the above to zero;

S1 = month when dynamic predictions are made in days / 640.5;

S2 = (month when dynamic predictions are made in days / 640.5)^2.

2. Using the above, XB = 0.0051*30 - 0.11*0 - 0.16*1 - 0.15*1 + 0.024*0 + (0.35*0 + 1.06*0 + 1.63*0 + 0.40*0) - 8.68*0 + 1.39*0

   XB = -0.1607

3. Therefore,

   S = 1 - Baseline survival for month 0^exp(XB)

   S = 1 - 0.6448 ^exp(-0.1607) = 0.31

   i.e. at workup completion couple A have a **31%** chance of conception over the following six months.

4. As a further example, if Couple A have IVF treatment six months after workup completion their chances of conception over the next six months is:

   XB = 0.0051*30 - 0.11*0 - 0.16*1 - 0.15*1 + 0.024*0 + (0.35*0 + 1.06*0 + 1.63*1+ 0.40*0) - 8.68*(183/640.5) + 1.39*(183/640.5)^2

   XB = -0.90115

   Therefore,

   S = 1 - Baseline survival for month 6^exp(XB)

   S = 1 - 0.06362 ^exp(-0.90115) = 0.67, i.e. if Couple A have IVF at six months after workup completion they have a **67%** chance of conception over the following six months.