Running title:

IVF for unexplained subfertility; whom should we treat?

van Eekelen R[1*,2], van Geloven N[3], van Wely M[1], Bhattacharya S[4], van der Veen F[1], Eijkemans MJ[2], McLernon DJ[5]

[1] Centre for Reproductive Medicine, Academic Medical Centre, Meibergdreef 9 1105 AZ Amsterdam, the Netherlands

[2] Department of Biostatistics and Research Support, Julius Centre, University Medical Centre Utrecht, Heidelberglaan 100, 3584 CX Utrecht, the Netherlands

[3] Medical Statistics, Department of Biomedical Data Sciences, Leiden University Medical Centre, Einthovenweg 20, 2333 ZC Leiden, the Netherlands

[4] Cardiff University School of Medicine, Heath Park Cardiff CF14 4XN, United Kingdom

[5] Medical Statistics Team, Institute of Applied Health Sciences, University of Aberdeen, AB24 2ZD Aberdeen, United Kingdom

* Correspondence address. Email: r.vaneekelen@amc.uva.nl

Extended abstract

Study question: Which couples with unexplained subfertility can expect increased chances of ongoing pregnancy with in vitro fertilisation (IVF) compared to expectant management?

Summary answer: IVF is associated with higher chances of conception than expectant management in couples in whom the female partner is under 40 years of age. In contrast, IVF is less effective in women aged over 40 and in couples with one year of secondary subfertility regardless of the age of the woman.

What is known already: The clinical indications for IVF have expanded over time from bilateral tubal blockage to include unexplained subfertility in which there is no identifiable barrier to conception. Yet, there is little evidence from randomised controlled trials that IVF is effective in these couples.

Study design, size, duration: We compared outcomes in British couples with unexplained subfertility undergoing IVF (n=40.921) from registry data to couples with the same type of subfertility on expectant management. The latter comprised a prospective nation-wide Dutch cohort (n=4.875) and a retrospective regional cohort from Aberdeen, Scotland (n=975). We excluded couples who had tried for less than a year to conceive and also those with anovulation, uni- or bilateral tubal occlusion, mild or severe endometriosis or male subfertility i.e. impaired semen quality according to WHO criteria.

Participants/materials, setting, methods:
We matched couples who received IVF and couples on expectant management based on their characteristics to control for confounding. We fitted a Cox proportional hazards model including patient characteristics, IVF treatment and their interactions to estimate the individualised chance of conception over one year – either following IVF or expectant

management for all combinations of patient characteristics. The endpoint was conception leading to ongoing pregnancy defined as a foetus reaching a gestational age of at least 12 weeks.

Main results and the role of chance: The adjusted one year chance of conception was 47.9% (95%CI: 45.0-50.9) after IVF and 26.1% (95%CI: 24.2-28.0) after expectant management. The absolute difference in the average adjusted one year chances of conception was 21.8% (95%CI: 18.3-25.3) in favour of IVF.
The effectiveness of IVF was influenced by female age, duration of subfertility and previous pregnancy. IVF was effective in women under 40 years, but the chance of an IVF conception over one year declined sharply in women over 34. In contrast, in woman over 40 years of age, IVF was less effective, with an absolute difference in chance compared to expectant management of 10% or lower. Regardless of female age, IVF was also less effective in couples with a short period of secondary subfertility (1 year), who had chances of natural conception of 30% or above.

Limitations, reasons for caution: The one year chances of conception were based on three cohorts with different sampling mechanisms. Despite adjustment for the three most important prognostic patient characteristics, namely female age, duration of subfertility and primary or secondary subfertility, our estimates might not be free from residual confounding.

Wider implications of the findings: IVF should be used selectively in those who have the most to gain from active treatment over expectant management. Our results can be used by clinicians to counsel couples with unexplained subfertility, to inform their expectations and facilitate evidence-based, shared decision making.

Keywords

In vitro fertilisation; unexplained subfertility; natural conception; expectant management; cohort

Introduction

Subfertility is defined as not conceiving within one year of regular unprotected intercourse and this affects approximately one in nine heterosexual couples (Datta *et al.*, 2016). Following standard investigations, no cause can be identified in a third of these couples who are said to have unexplained subfertility. In vitro fertilisation (IVF), with or without intra-cytoplasmic sperm injection (ICSI), is a commonly used treatment for couples with prolonged unresolved subfertility and over 470.000 treatment cycles were recorded in Europe in 2013 (Calhaz-Jorge *et al.*, 2017). IVF is a burden to couples in terms of mental and physical stress, is associated with high expectations and considerable investment in terms of emotions, finances and time (Rooney and Domar, 2016). The number of IVF cycles conducted increases annually, posing an increasing burden on health services in countries where IVF is publicly funded (HFEA, 2004; Andersen *et al.*, 2007; NVOG, 2010; NICE, 2013; Kamphuis *et al.*, 2014; Calhaz-Jorge *et al.*, 2017; HFEA, 2018). This increase is generally considered to be the consequence of the increasingly liberal utilisation of IVF for a variety of indications including unexplained subfertility (HFEA, 2004; Kamphuis *et al.*, 2014; HFEA, 2015). Yet, there is little robust evidence supporting the effectiveness of IVF in couples with unexplained subfertility compared to a wait-and-see approach i.e. expectant management (Pandian *et al.*, 2015; Tjon-Kon-Fat *et al.*, 2016).

There is a single trial evaluating the effectiveness of IVF versus expectant management for couples with unexplained subfertility in terms of live birth which reported the chance of live birth following IVF (11 out of 24 couples) to be12 times that of expectant management (1 out of 27 couples) (Hughes *et al.*, 2004). Although the results seem to support IVF, there is considerable uncertainty around this result based on very small numbers of participants and it is inappropriate for clinical practice across the globe to be based on this quality of evidence (Tjon-Kon-Fat *et al.*, 2016).

Observational studies have separately quantified the predicted chances of conception after IVF and after a period of expectant management (Leushuis *et al.*, 2009; McLernon *et al.*, 2016; van Eekelen *et al.*, 2017a). There are two problems that hamper the comparability

of these predictions which currently limit their clinical utility. First, the prognoses were derived from separate studies with dissimilar patient characteristics. For instance, women with unexplained subfertility who received IVF are generally older than women who pursued expectant management. Second, the prognosis after IVF is expressed per embryo transfer or per complete IVF cycle while the prognosis associated with expectant management is expressed in terms of calendar time, commonly over one year (Daya, 2005).

We can address these problems by adjusting for differences between couples who were treated with IVF and couples who pursued expectant management and expressing predicted chances over a uniform time horizon. To this end, we opted for a pragmatic approach by analysing data from three observational cohorts: the UK national IVF registry and two groups of couples (from the Netherlands and Scotland respectively) who embarked on a variable period of expectant management.

Our aim was threefold: first to use individual patient data from these three cohorts to compare the average absolute unadjusted adjusted one year chance of conception after IVF or expectant management, second to compare the adjusted one year chance of conception after IVF or expectant management and third, to estimate the effectiveness of IVF in individual patients based on their clinical characteristics.

Materials and Methods

In short: the population comprised couples with unexplained subfertility seen in fertility clinics. The exposure was all IVF cycles and subsequent embryo transfers received within one year after the start of ovarian stimulation. The comparator in the unexposed group was expectant management for one year after completion of the fertility workup. The outcome of interest was conception leading to ongoing pregnancy.

IVF cohort

Data on couples treated with IVF between 1999 and 2011 were obtained from the Human Fertilisation and Embryology Authority (HFEA) registry which collects data from all licensed clinics in the United Kingdom (McLernon *et al.*, 2016). From 2009 onwards, the number of women included was limited because explicit consent was required for the use of their data for research purposes (McLernon *et al.*, 2016).

Expectant management cohorts

We combined data from two separate cohorts comprising couples with unexplained subfertility who underwent expectant management. The first was a prospective cohort assembled across 38 hospitals in The Netherlands between January 2002 and February 2004. Couples were followed for natural conception from the completion of the fertility workup onwards. The detailed protocol for this has been described elsewhere (van der Steeg *et al.*, 2007). The second was a retrospective population based cohort from the Grampian region of Scotland comprising subfertile couples who registered at Aberdeen Fertility Clinic. Using a unique, pseudonomised identifier, we linked patient records including demographic and diagnostic information from the fertility clinic to treatment records from Aberdeen Assisted Reproduction Unit Database and to pregnancy outcomes from the Aberdeen Maternity and Neonatal Databank (van Eekelen *et al.*, 2018). This process was carried out according to the Standard Operating Procedures of the Data Management Team, University of Aberdeen. We selected couples living in the Aberdeen City District whose births occurred at Aberdeen Fertility Clinic. Pregnancy outcomes from natural conceptions were identified by linkage with the Aberdeen Maternity and Neonatal Databank which captures all birth outcomes in this region (Ayorinde *et al.*, 2016).

Inclusion and exclusion criteria

Couples who had been trying for a pregnancy for less than one year, those with anovulation, uni- or bilateral tubal occlusion, mild or severe endometriosis and male subfertility i.e. impaired semen quality according to WHO criteria were excluded from the UK IVF and

Scottish cohorts (WHO, 1999; WHO, 2010). For the Dutch cohort, the same exclusion criteria were applied, except that mild endometriosis was considered as a part of unexplained subfertility and male subfertility was defined as a total motile count below 1 million (van Eekelen *et al.*, 2017a).

Treatment protocols

Decisions regarding treatment were based on local and national protocols. In short, the UK IVF registry comprises every IVF cycle with guidelines changing over time (NICE, 2013). Treatment decisions for the Dutch cohort were left to the discretion of physicians in agreement with their patients (NVOG, 2004; van der Steeg *et al.*, 2007) and in the Scottish cohort by the local protocol and national guideline (NICE, 2013).

Expectant management was defined as no intervention aside from the advice to have intercourse.

Definitions for outcome and follow up

Our outcome of interest was conception leading to an ongoing pregnancy, defined as a foetus reaching a gestational age of at least 12 weeks visualised by ultrasound. The date of conception was defined as the first day of the last menstruation period prior to conception. We analysed data up to a maximum of one year of follow up.

Follow up for couples on expectant management started at completion of the fertility workup and ended, for those who did not conceive, at one year after the workup, on the date of last contact or the date of starting ovarian stimulation for IUI or IVF treatment (whichever came first) i.e. we censored their time-to-pregnancy. We assumed that couples who continued with expectant management were no different, in terms of their clinical characteristics and resulting prognosis, to those who were censored (non-informative censoring).

Couples who received IVF were followed from the start of ovarian stimulation in the first cycle up until their last embryo transfer. Since the IVF registry contained all UK IVF cycles from

1999 to 2011, all ongoing IVF pregnancies within a year of initiating the first cycle (i.e. all fresh and frozen cycles) were recorded and we thus had complete one year follow up during which couples received 1.5 embryo transfers on average. This assumes that couples who discontinued treatment had zero chance of conception after IVF afterwards, for instance for reasons related to an insufficient number of oocytes collected during follicle aspiration, a low fertilization rate or financial reasons (Daya, 2005).

To align with our assumption of pursuing one full year of expectant management, we also considered the hypothetical scenario in which couples continued their IVF attempts for a full year of follow up during which they underwent 3 to 4 embryo transfers on average. In the supplementary analysis following this scenario, we censored time-to-pregnancy in couples receiving IVF after their last unsuccessful IVF transfer, defined as the first day of menstruation before the last embryo transfer. We thus also assumed non-informative censoring in IVF i.e. that couples who continued IVF were similar to couples who dropped out of IVF.

Missing data

To be able to compare couples who received IVF and couples who expectant management, we had to make assumptions around the dates of ovarian stimulation and first day of menstruation in couples who had IVF. As couples start their IVF treatment with ovarian stimulation, we elected to follow couples from that date until conception (the first day of last menstruation before the final embryo transfer) to align with the general definition of time to natural conception. Since dates of initiation of ovarian stimulation were not available in the UK IVF database and are not applicable to frozen/thawed cycles, we assumed a period of 15 days before the date of embryo transfer (Alport *et al.*, 2011).

In the Dutch cohort, the date of workup completion could be derived and this date was used as the start of follow up (van Eekelen *et al.*, 2017a). For the Scottish cohort, this date was not available and was estimated at six weeks after the date of registration, which was the average time between registration and completion of the fertility workup in the Dutch cohort.

The prognostic patient characteristics that were recorded in all cohorts were female age, duration of subfertility and (female) primary or secondary subfertility. In the UK IVF cohort, data for primary or secondary subfertility from 2008 onwards (n=7532, 18%) were not systematically recorded and were considered as missing. Because of these missing values, we applied multiple imputation including all relevant prognostic characteristics and a covariate for the cumulative hazard of pregnancy to account for the aspect of time in the data, creating 10 imputation sets (White and Royston, 2009). In the Dutch cohort, fewer than 1% of data used for the present study were missing and were accounted for in a previous study by multiple imputation, creating 10 imputation sets (van Eekelen *et al.*, 2017a). In the Scottish cohort, fewer than 1% of data were missing and we applied multiple imputation identical to the approach in the UK IVF cohort. Ten imputation sets were thus created separately for the three cohorts, then combined to derive 10 combined datasets and we pooled their results using Rubin's Rules (Rubin, 2004).

Matching procedure

To ensure that there was no confounding due to the three prognostic patient characteristics (female age, duration of subfertility and previous pregnancy), we applied matching (Austin, 2014). In this matching procedure, we paired couples on expectant management to couples that received IVF that had the same (rounded) female age, duration of subfertility and primary or secondary subfertility status. We found all possible pairs with replacement which allows each patient to be used as a match more than once. This yields higher quality matches than matching without replacement due to data on all matches being used (Abadie and Imbens, 2006). Then, we weighted couples such that the expectant management group was the reference or 'target population'. Thus, in the resulting complete 'matched' dataset, the average patient characteristics and sample size of couples on expectant management were now identical to couples who received IVF. Using this matched data, we estimate what would happen if couples on expectant management would instead start IVF (referred to as the average treatment effect in controls, or ATC) (Austin, 2014).

Statistical analysis

Average effect of IVF

We calculated the unadjusted one year chance of conception after IVF as the observed fraction of couples who conceived within one year of IVF on the original, unmatched dataset. We estimated the unadjusted one year chance of conception after expectant management with the Kaplan-Meier method on the original, unmatched dataset. We calculated the average unadjusted effect as the absolute difference of these two chances. To estimate the adjusted chances and the adjusted average effect, we repeated both these analyses on the matched dataset.

Individualised effectiveness of IVF

We defined the individualised effectiveness of IVF as the absolute difference between the estimated one year chance of conception after IVF and the one year chance when pursuing expectant management for a couple based on female age, duration of subfertility and primary/secondary subfertility status. To estimate these individual chances, we fitted a Cox proportional hazards model on the original, unmatched dataset using treatment (IVF or expectant management), the patient characteristics and the interaction between treatment and patient characteristics as covariates. This was done following three steps.

We first determined how female age and duration of subfertility could best be entered into our statistical model: we evaluated both linear and non-linear associations with the log hazard of conception using linear terms or restricted cubic splines, then tested which fitted better using Wald tests and Akaike's Information Criterion (AIC) (Akaike, 1974; Harrell *et al.*, 1996).

Once a suitable form for female age and duration of subfertility was determined, we included IVF treatment, female age, duration of subfertility, primary or secondary subfertility and all interaction terms with IVF treatment in the model to assess if the effect of IVF depended on these characteristics. We then tested all interaction terms simultaneously with

an overall Wald test. If this test was significant, we performed backwards selection on the full model using Wald tests per separate interaction and AIC to determine which interaction was informative and removed those that were not (Akaike, 1974). We checked the proportional hazards assumption for all covariates in the model using scaled Schoenfeld residuals (Grambsch and Therneau, 1994) and accounted for the non-proportional hazard for IVF treatment versus expectant management by stratifying on treatment group.

After the final model fit, we visualized the association between patient characteristics which varied the effect of IVF by estimating one year chances of conception for couples with different characteristics.

In addition, we estimated chances for all combinations of patient characteristics, tabulating the estimated chances, their corresponding 95% confidence intervals (CIs), absolute differences, relative differences and the number needed to treat (NNT).

Supplementary analyses

In the first supplementary analysis, in order to estimate the outcome if couples would continue to have IVF over a full one year, we used the Kaplan-Meier method both for couples receiving IVF and for couples pursuing expectant management on the original and matched datasets.

In the second supplementary analysis, we again estimated individualised chances after both IVF and expectant management but now expressed over a period of 6 months. We tabulated these 6 month chances as well as their corresponding 95% confidence intervals (CIs), absolute differences, relative differences and the number needed to treat (NNT).

The study was approved by the North of Scotland Research Ethics Committee (17/NS/0122). Data linkage and all statistical analyses were performed in the Data Safe Haven of the University of Aberdeen using R version 3.4.3 (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/) and RStudio using the *survival* package for the Kaplan-

Meier method, *mice* for multiple imputation of missing data, *rms* for functions for splines and fitting Cox models and *Matching* to conduct the matching by patient characteristics.

Results

Data from a total of 46.771 couples were available for analysis (**Figure 1**). Out of 40.921 couples in the UK IVF cohort who received 61.019 embryo transfers in total, 16.281 conceived (39.8% of couples, 26.7% per embryo transfer) within one year of starting IVF. 32.396 (79%) couples received IVF and 8.525 (21%) received ICSI. There were 4.891 multiple gestations after IVF (12% of couples, 30% of conceptions). Out of 4.875 couples in the Dutch cohort pursuing expectant management, 903 (18.5%) couples conceived naturally within one year after completion of the fertility workup. There were 11 multiple gestations (0.2% of couples, 1.2% of conceptions). Out of 975 couples in the Scottish cohort pursuing expectant management, 229 (23.5%) couples conceived naturally within one year after completion of the fertility workup. There were no multiple gestations.

The median duration of follow up for couples receiving IVF was one embryo transfer (25th-75th percentile: 0-7 months) as 29% of couples conceived after their first embryo transfer and 21% discontinued IVF treatment after their first unsuccessful embryo transfer. The median follow up for couples pursuing expectant management was 7 months (25th-75th percentile: 3-12 months).

Patient characteristics

The baseline characteristics of couples, stratified by cohort, are presented in **Table I**. In comparison with women who were managed expectantly, those who received IVF were older (mean 35.1 years in the UK IVF, 32.5 years in the Dutch and 33.2 years in the Scottish cohorts), had been trying to conceive for longer (median 4.0 years in UK IVF, 1.6 years in the Dutch and 2.1 years in the Scottish cohorts) but were just as likely to have primary subfertility (60% in the UK IVF, 66% in the Dutch and 59% in the Scottish cohorts).

The distributions of female age and duration of subfertility for couples who received IVF and couples who pursued expectant management are shown in **Figures 2A** and **B**.

Unadjusted average chance of conception

The unadjusted one year chance of conception after starting IVF was 39.8% (95%CI: 39.3-40.3) and after expectant management was 26.1% (95%CI: 24.7-27.5). The average absolute difference in the unadjusted one year chance of conception was 13.6% (95%CI: 11.6-15.7) in favour of IVF. The one year chances following expectant management in the Dutch and Scottish cohorts were similar (26.9% and 23.8% respectively).

Adjusted average chance of conception

A total of 5.818 out of 5.850 (99%) couples pursuing expectant management were matched with 31.867 out of 40.921 (78%) counterparts who received IVF and had the same characteristics. The adjusted one year chance of conception was 47.9% (95%CI: 45.0-50.9) after starting IVF and 26.1% (95%CI: 24.2-28.0) after expectant management. The average absolute difference in the adjusted one year chance of conception was 21.8% (95%CI: 18.3-25.3) in favour of IVF.

Individualised effectiveness of IVF

Both female age and duration of subfertility were non-linearly associated with conception (Wald tests for non-linearity both $p<0.001$, splines with 5 and 3 knots respectively).
There were statistically significant interactions between all three patient characteristics and IVF treatment (overall $p<0.001$, individual interactions all $p<0.001$)**.**
The estimated effects of couple characteristics on conception in terms of hazard ratios (HRs) are presented in **Table II.** In general, as female age increased, the chance of conception decreased both after expectant management and after IVF, but the detrimental effect of female age above 34 years on the chance of conception was stronger in the latter (HR of 40 versus 35 years: 0.43 after IVF and 0.64 after expectant management). As duration of

subfertility increased, the chance of conception decreased in both groups, but this effect was stronger for those on expectant management (HR of 6 versus 2 years: 0.86 after IVF and 0.39 after expectant management). Couples with primary subfertility on expectant management had a lower chance of conception compared to couples with secondary subfertility (HR of primary versus secondary: 0.71) but there was no noticeable difference in the IVF group (HR: 0.98).

The predicted one year chance of conception in couples with primary subfertility of 2 years duration and female age ranging between 26 and 42 are shown in **Figure 3**. The effectiveness of IVF decreased in women over 34 years.

The predicted one year chances of conception in couples with primary subfertility where female age is 35 years and the duration of subfertility ranges from 1 to 8 years are visualised in **Figure 4**. The effectiveness of IVF increased as the duration of subfertility increased.

The predicted one year chances of conception for couples with 2 year duration where female age is 35 years stratified for primary and secondary subfertility are presented in **Table III**. IVF was more effective for couples with primary subfertility than for couples with secondary subfertility.

In **Supplementary Material I**, we present full tables containing the predicted one year chance of conception after IVF and after starting expectant management for all combinations of patient characteristics. Also provided are the absolute differences between these chances, the relative differences and the numbers needed to treat (NNT) to achieve one additional conception.

For instance, a typical couple undergoing IVF, where the woman is 35 years old with 4 years duration of primary subfertility, has an estimated one year chance of conception of 46% (95%CI: 44-48) after IVF compared to 12% (95%CI: 9-14) after expectant management, with an absolute difference of 34% and a NNT of 2.9.

On the other hand, a typical couple pursuing expectant management, where the woman is 33 years old with 2 years of primary subfertility, has an estimated one year chance of

conception of 53% (95%CI: 50-55) after IVF compared to 23% (95%CI: 20-25) after expectant management, with an absolute difference of 30% and a NNT of 3.3.

In couples where the woman is under 40 years, IVF was effective compared to expectant management. In contrast, in couples where the woman is over 40 years, IVF was less effective as the absolute difference between chances was approximately 10% or lower. In couples with one year duration of secondary subfertility, regardless of the age of the woman, IVF was also less effective since their chances of natural conception remained relatively high at 30% or above.

Supplementary analyses

In the supplementary analysis where we estimated outcomes in couples who continued with IVF for a full year, the unadjusted one year chance of conception after IVF was estimated at 51.6% (95%CI: 50.9-52.2). The average absolute difference in the unadjusted one year chance of conception became 25.4% (95%CI: 23.1-27.7) in favour of IVF.

The adjusted one year chance of conception after receiving IVF for one full year was estimated at 59.7% (95%CI: 55.3-64.0). The average absolute difference in the adjusted one year chance of conception became 33.6% (95%CI: 28.8-38.3) in favour of IVF.

In **Supplementary Material II**, we present the same individualised predictions as in **Supplementary Material I** but now expressed over 6 months instead of one year.

Discussion

In couples with unexplained subfertility, we found that IVF increased the average one year chance of conception compared to expectant management. Factors affecting the effectiveness of IVF were female age, duration of subfertility and primary/secondary subfertility.

Although couples who received IVF had, on average, a higher female age and a higher duration of subfertility compared to couples who continued expectant management, the large sample size of treated and untreated couples resulted in sufficient overlap of case-mix to enable us to accurately estimate all the separate interactions between patient characteristics and treatment. A second strength was our ability to control for confounding in the average adjusted chance by matching on female age, duration of subfertility and primary versus secondary subfertility.

We were able to predict individualised chances of conception following either IVF or expectant management on the same time axis representing one year of 'real' calendar time. This is intuitive, allows for a straightforward comparison, allows for most couples to complete at least one full IVF cycle and is easier to communicate to patients compared to chances per embryo transfer or per IVF cycle. A longer follow up might increase the rates after both IVF and expectant management but may be more difficult for decision making, as the longer the follow up period becomes, the less likely couples are to continue IVF.

Aside from calculating the observed fraction of couples who conceived within one year in the matched data (approximately 48%), we also estimated the adjusted chance of conception when receiving IVF for one full year i.e. when continuing IVF (approximately 60%). The latter might be an optimistic estimate, as not all couples can continue with additional IVF cycles, for instance because of an insufficient number of oocytes or financial reasons.

Limitations of this study include the availability of only three important patient characteristics in all data sources, the missing date of completion of the fertility workup in the Scottish data and the possibility of residual confounding due to the observational nature of the data. We had to make an assumption on the time between registration and completion of the fertility workup in the Scottish cohort. In the Dutch cohort, this was on average six weeks (van Eekelen *et al*., 2019). In a previously conducted validation study, we found similar chances of ongoing pregnancy in the Scottish and Dutch cohort when assuming six weeks between

registration and completion of the fertility workup, hence this assumption was deemed reasonable (van Eekelen *et al*., 2019). The dropout rate after the first embryo transfer of 21% is higher than the 12% reported in a recent Dutch validation study, but the difference can be explained by the geographical variation in reimbursement for the UK IVF cohort compared to full reimbursement up to three cycles at the time of the Dutch study (Leijdekkers *et al*., 2018). In addition, the three different data sources used different sampling mechanisms, which could potentially compromise the comparability of study populations. Couples pursuing expectant management were recruited at completion of the fertility workup (Dutch cohort) or identified retrospectively (Scottish cohort). In contrast, couples who received IVF were registered in the UK IVF database with no prior data other than diagnosis. Therefore we were unable to assess or adjust for any selection bias that might occur between completion of the fertility workup and the start of treatment, as only couples that did not conceive naturally during that period will have ended up in the UK IVF registry, a selection which might not be fully captured by the duration of subfertility (van Eekelen *et al*., 2017b).

As the UK IVF data were only available up to 2011 and treatment success rates were found to increase over time, our estimates for the one year chance after IVF might be conservative for today's practice. However, IVF rates in the UK in 2016 were found to plateau in 2013 to 25%-26% per cycle (HFEA, 2016; HFEA, 2018). A recent external validation of the OPiS model developed on UK IVF data up to 2008 showed good performance in Dutch data collected up to 2014, meaning that our data might reasonably reflect today's practice and pregnancy outcomes (McLernon *et al.*, 2016; Leijdekkers *et al.*, 2018). The decade has witnessed changes in embryo transfer protocols in the UK from predominantly double embryo transfer (DET) to increasing numbers of elective single embryo transfer (eSET) resulting in a decline in multiple pregnancy rates from 27% in 2008 to16% in 2014 (Harbottle *et al.*, 2015; HFEA, 2015). Nevertheless, the impact of this change in IVF policy on our estimated chances of conception might be minor as the cumulative chances of IVF success are comparable following DET and eSET combined with subsequent transfers

of frozen/thawed embryos (Lukassen *et al.*, 2005; McLernon *et al.*, 2010; Harbottle *et al.*, 2015).

The primary outcome was ongoing pregnancy because the increased logistical efforts and associated costs involved in following couples to delivery was not possible in the Dutch cohort. Ongoing pregnancy is generally considered an appropriate proxy for live birth in clinical research: approximately 95% of ongoing pregnancies lead to live birth (Clarke *et al.*, 2010; Braakhekke *et al.*, 2014).

A large RCT would be the ideal study design to assess the effectiveness of IVF compared to expectant management. Conducting such a trial now would be challenging as IVF has become an established treatment for unexplained subfertility and many couples are unconfident about the value of expectant management, overestimate IVF success and push for early active treatment (van den Boogaard *et al.*, 2011; Kersten *et al.*, 2015). In addition, many clinicians fail to take into account couples' chances of natural conception in their consultations and believe that it would be unethical to withhold early access to IVF (Kersten *et al.*, 2015). This has created a genuine lack of equipoise without which no trial can be conducted. We therefore felt that the best and most pragmatic option was to compare observational data from cohorts on expectant management and IVF (van Eekelen *et al.*, 2017b).

A key benefit of the present study is the provision of the adjusted average effectiveness of IVF compared to expectant management and, in addition, individualised estimates which are easy to interpret and allow for direct comparisons.

Our results may be used by clinicians to counsel couples with unexplained subfertility to inform their expectations and to avoid unnecessary treatment for some whilst allowing timely access to IVF for others. They can also be used to allow funders and commissioners to make decisions on access to publicly funded IVF.

Our results need to be validated in other datasets or, ideally, in RCTs involving couples with characteristics in whom the effectiveness of IVF is unclear and some equipoise remains. In

addition, data on long term follow up after the first live birth is necessary to counsel couples who wish to have multiple children.

Conclusion

The effectiveness of IVF over expectant management in unexplained subfertility depends on the characteristics of the couple. IVF should be used selectively in those who have the most to gain from active treatment over expectant management. Our results can be used by clinicians to counsel couples with unexplained subfertility, to inform their expectations and facilitate evidence-based, shared decision making.

Author's roles:

NvG, MDJ, BS, FvdV, MvW and MJE conceived the study. MDJ oversaw the storage of all data in the Safe Haven. RvE performed the data linkage. RvE, NvG, MJE and MDJ designed the statistical analysis plan. RvE and MDJ analysed the data. RvE, NvG, FvdV and BS drafted the manuscript. All authors contributed critical revision to the paper and approved the final manuscript.

Funding

This work was supported by Tenovus Scotland, grant G17.04.

Travel for RvE was supported by the Amsterdam Reproduction & Development Research Group, grant V.000296.

Conflicts of interest

BS reports acting as Editor-in-Chief of HROpen.

Acknowledgements

The authors would like to thank Tenovus Scotland and the Amsterdam Reproduction & Development Research Group for funding this project.

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.

References

Abadie A and Imbens GW. Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica* 2006;**74**:235-267.

Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974;**19**:716-723.

Alport B, Case A, Lim H and Baerwald A. Does the Ovarian Stimulation Phase Length Predict In vitro Fertilization Outcomes? *Int J Fertil Steril* 2011;**5**:134-141.

Andersen AN, Goossens V, Gianaroli L, Felberbaum R, de Mouzon J and Nygren KG. Assisted reproductive technology in Europe, 2003. Results generated from European registers by ESHRE. *Hum Reprod* 2007;**22**:1513-1525.

Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Stat Med* 2014;**33**:1242-1258.

Ayorinde AA, Wilde K, Lemon J, Campbell D and Bhattacharya S. Data Resource Profile: The Aberdeen Maternity and Neonatal Databank (AMND). *Int J Epidemiol* 2016;**45**:389-394.

Braakhekke M, Kamphuis EI, Dancet EA, Mol F, van der Veen F and Mol BW. Ongoing pregnancy qualifies best as the primary outcome measure of choice in trials in reproductive medicine: an opinion paper. *Fertil Steril* 2014;**101**:1203-1204.

551 Calhaz-Jorge C, De Geyter C, Kupka MS, de Mouzon J, Erb K, Mocanu E, Motrenko T,
552 Scaravelli G, Wyns C and Goossens V. Assisted reproductive technology in Europe, 2013:
553 results generated from European registers by ESHRE. *Hum Reprod* 2017;**32**:1957-1973.

554

555 Clarke JF, van Rumste MM, Farquhar CM, Johnson NP, Mol BW and Herbison P. Measuring
556 outcomes in fertility trials: can we rely on clinical pregnancy rates? *Fertil Steril* 2010;**94**:1647-
557 1651.

558

559 Datta J, Palmer MJ, Tanton C, Gibson LJ, Jones KG, Macdowall W, Glasier A, Sonnenberg
560 P, Field N, Mercer CH*, et al.* Prevalence of infertility and help seeking among 15 000 women
561 and men. *Hum Reprod* 2016;**31**:2108-2118.

562

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

565

566 Grambsch PM and Therneau TM. Proportional hazards tests and diagnostics based on
567 weighted residuals. *Biometrika* 1994;**81**:515-526.

568

569 Harbottle S, Hughes C, Cutting R, Roberts S and Brison D. Elective Single Embryo Transfer:
570 an update to UK Best Practice Guidelines. *Hum Fertil (Camb)* 2015;**18**:165-183.

571

572 Harrell FE, Jr., Lee KL and Mark DB. Multivariable prognostic models: issues in developing
573 models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat*
574 *Med* 1996;**15**:361-387.

575

576 HFEA, Human Fertilisation and Embryology Authority. Annual report and accounts 2004/05.
577 Williams Lea Group, UK. 2004;

578

HFEA, Human Fertilisation and Embryology Authority. Annual report and accounts 2015/16. Williams Lea Group, UK.

HFEA, Human Fertilisation and Embryology Authority. Annual report and accounts 2016/17. Williams Lea Group, UK.

HFEA, Human Fertilisation and Embryology Authority. Fertility treatment 2014-2016: Trends and figures**:** 60.

Hughes EG, Beecroft ML, Wilkie V, Burville L, Claman P, Tummon I, Greenblatt E, Fluker M and Thorpe K. A multicentre randomized controlled trial of expectant management versus IVF in women with Fallopian tube patency. *Hum Reprod* 2004;**19**:1105-1109.

Kamphuis EI, Bhattacharya S, van der Veen F, Mol BW and Templeton A. Are we overusing IVF? *BMJ* 2014;**348**:g252.

Kersten FA, Hermens RP, Braat DD, Hoek A, Mol BW, Goddijn M and Nelen WL. Overtreatment in couples with unexplained infertility. *Hum Reprod* 2015;**30**:71-80.

Leijdekkers JA, Eijkemans MJC, van Tilborg TC, Oudshoorn SC, McLernon DJ, Bhattacharya S, Mol BWJ, Broekmans FJM and Torrance HL. Predicting the cumulative chance of live birth over multiple complete cycles of in vitro fertilization: an external validation study. *Hum Reprod* 2018;**33**:1684-1695.

Leushuis E, van der Steeg JW, Steures P, Bossuyt PM, Eijkemans MJ, van der Veen F, Mol BW and Hompes PG. Prediction models in reproductive medicine: a critical appraisal. *Hum Reprod Update* 2009;**15**:537-552.

Lukassen HG, Braat DD, Wetzels AM, Zielhuis GA, Adang EM, Scheenjes E and Kremer JA. Two cycles with single embryo transfer versus one cycle with double embryo transfer: a randomized controlled trial. *Hum Reprod* 2005;**20**:702-708.

McLernon DJ, Harrild K, Bergh C, Davies MJ, de Neubourg D, Dumoulin JC, Gerris J, Kremer JA, Martikainen H, Mol BW*, et al.* Clinical effectiveness of elective single versus double embryo transfer: meta-analysis of individual patient data from randomised trials. *Bmj* 2010;**341**:c6945.

McLernon DJ, Steyerberg EW, Te Velde ER, Lee AJ and 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* 2016;**355**:i5735.

NICE, National Institute for Health and Care Excellence. Guideline on: Fertility problems: assessment and treatment (2013). Accessed on: 5th of February, 2017. Available from: https://www.nice.org.uk/guidance/cg156.

NVOG, Dutch Society for Obstetrics and Gynaecology. Guideline on basic fertility workup (2004).

NVOG, Dutch Society for Obstetrics and Gynaecology. Guideline on: subfertility (2010). Accessed on: 5th of February, 2017. Available from: http://bit.ly/1UhuYMV.

Pandian Z, Gibreel A and Bhattacharya S. In vitro fertilisation for unexplained subfertility. *Cochrane Database Syst Rev* 2015;**2**:Cd003357.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

636 Rooney KL and Domar AD. The impact of stress on fertility treatment. *Curr Opin Obstet*
637 *Gynecol* 2016;**28**:198-201.

638

639 Rubin D (2004). Multiple Imputation for Nonresponse in Surveys. New York, John Wiley and
640 Sons.

641

642 Tjon-Kon-Fat RI, Bensdorp AJ, Scholten I, Repping S, van Wely M, Mol BW and van der
643 Veen F. IUI and IVF for unexplained subfertility: where did we go wrong? *Hum Reprod*
644 2016;**31**:2665-2667.

645

646 van den Boogaard NM, van den Boogaard E, Bokslag A, van Zwieten MC, Hompes PG,
647 Bhattacharya S, Nelen W, van der Veen F and Mol BW. Patients' and professionals' barriers
648 and facilitators of tailored expectant management in subfertile couples with a good prognosis
649 of a natural conception. *Hum Reprod* 2011;**26**:2122-2128.

650

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

654

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

659

660 van Eekelen R, McLernon DJ, van Wely M, Eijkemans MJ, Bhattacharya S, van der Veen F
661 and van Geloven N. External validation of a dynamic prediction model for repeated
662 predictions of natural conception over time. *Hum Reprod* 2018;**33**:2268-2275.

663

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

White IR and Royston P. Imputing missing covariate values for the Cox model. *Stat Med* 2009;**28**:1982-1998.

WHO, World Health Organisation. Laboratory Manual for the Examination of Human Semen and Sperm-Cervical Mucus Interaction. 4th edn 1999. Cambridge University Press, Cambridge.

WHO, World Health Organisation. Laboratory Manual for the Examination and Processing of Human Semen, 5th ed 2010. Geneva: World Health Organization.

Table I. Baseline characteristics at the start of follow up for the three cohorts included in the analysis.

	UK IVF (n=40.921)	Dutch (n=4.875)	Scottish (n=975)
Female age (mean, 5th-95th percentile)	35.1 (28-42)	32.5 (24.9-39.4)	33.2 (26.1-41.1)
Duration of subfertility (median, 5th-95th percentile)	4.0 (1-13)	1.6 (1-4.9)	2.1 (1.1-5.1)
Primary subfertility (n, %)	24572 (60%)	3231 (66%)	571 (59%)

Table II. Estimated effects of patient characteristics on conception leading to ongoing pregnancy. Results are from the model including interaction (via stratification) with treatment.

	Hazard ratio for conception after IVF (95%CI)	Hazard ratio for conception after expectant management (95%CI)
Female age, years (34 versus 27)*	0.99 (0.94-1.04)	0.70 (0.60-0.82)
Female age, years (40 versus 35)*	0.43 (0.41-0.46)	0.64 (0.49-0.84)
Duration of subfertility, years (6 versus 2)*	0.86 (0.80-0.92)	0.39 (0.30-0.50)
Primary versus secondary subfertility	0.98 (0.94-1.02)	0.71 (0.63-0.81)

*Contrasts between values for female age and duration of subfertility were chosen to depict their non-linear estimated effects.

Table III. Association between primary or secondary subfertility and the one year chance of conception after receiving IVF or pursuing expectant management for a couple of which the woman is 35 years old who have been trying to conceive for 2 years

	One year chance of conception after IVF (95%CI)	One year chance of conception after expectant management (95%CI)
Primary subfertile couple	49.2 (46.3-52.1)	19.9 (16.7-23.1)
Secondary subfertile couple	50.0 (47.0-53.0)	26.7 (22.2-31.2)

Figure 1. Flowchart of recruitment and inclusion/exclusion in the three cohorts

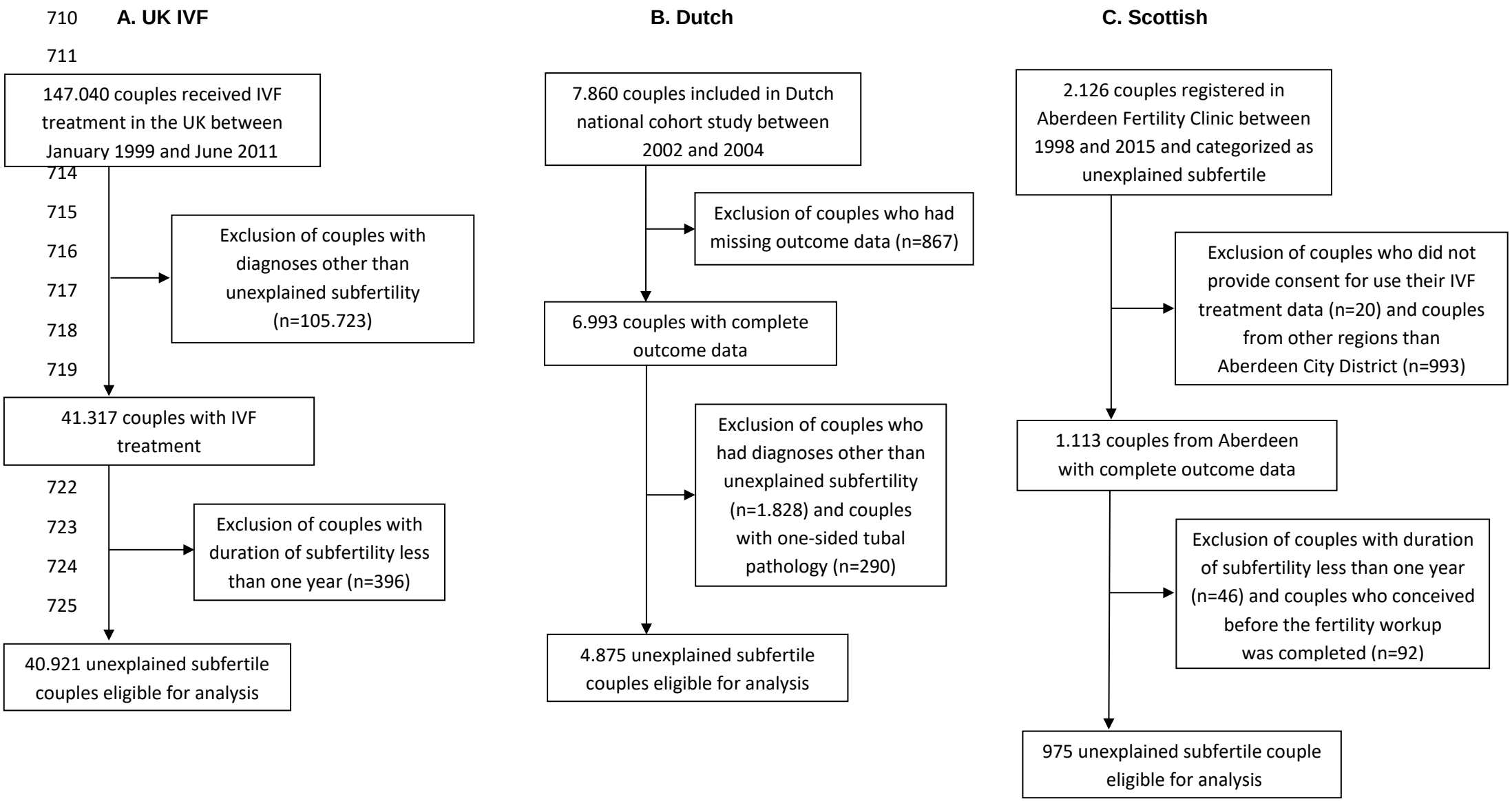
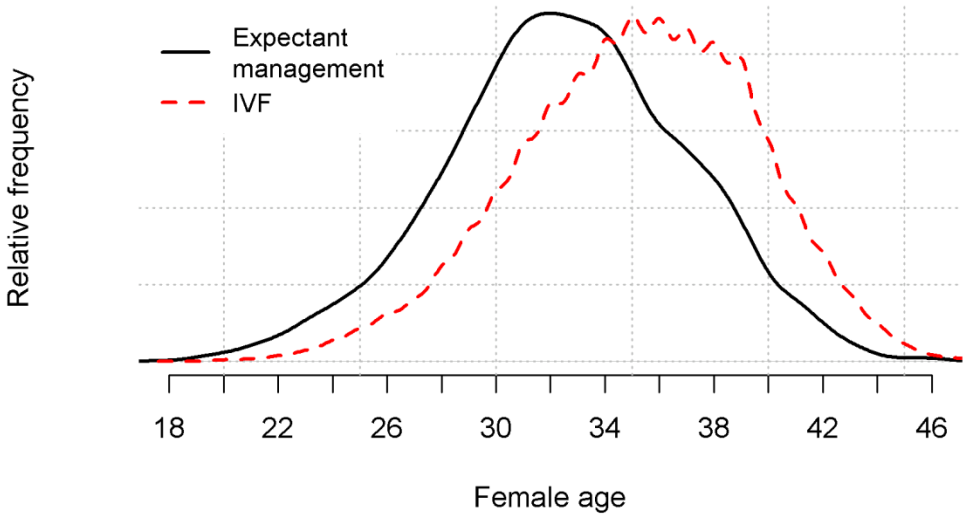


Figure 2. Overlap of patient characteristics for couples who received IVF and couples who pursued expectant management

A. Distribution of female age per treatment group, depicted by relative frequency (density)



B. Distribution of duration of subfertility per treatment group, depicted as the proportion of couples per group who had a certain (rounded) duration

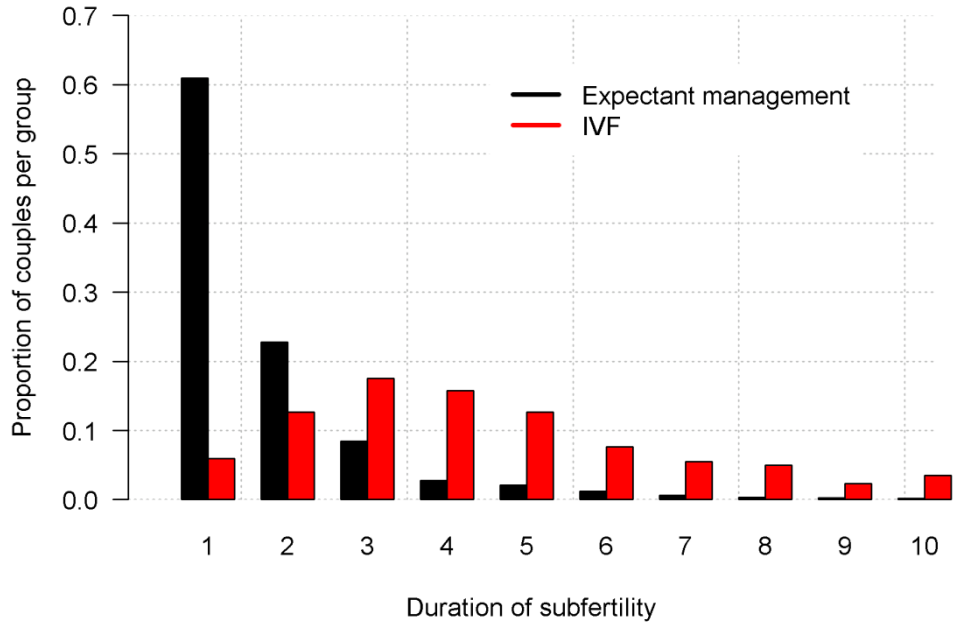


Figure 3. Association between female age and the one year chance of conception after receiving IVF or pursuing expectant management for a primary subfertile couple who have been trying to conceive for 2 years. Grey bands are 95% confidence limits

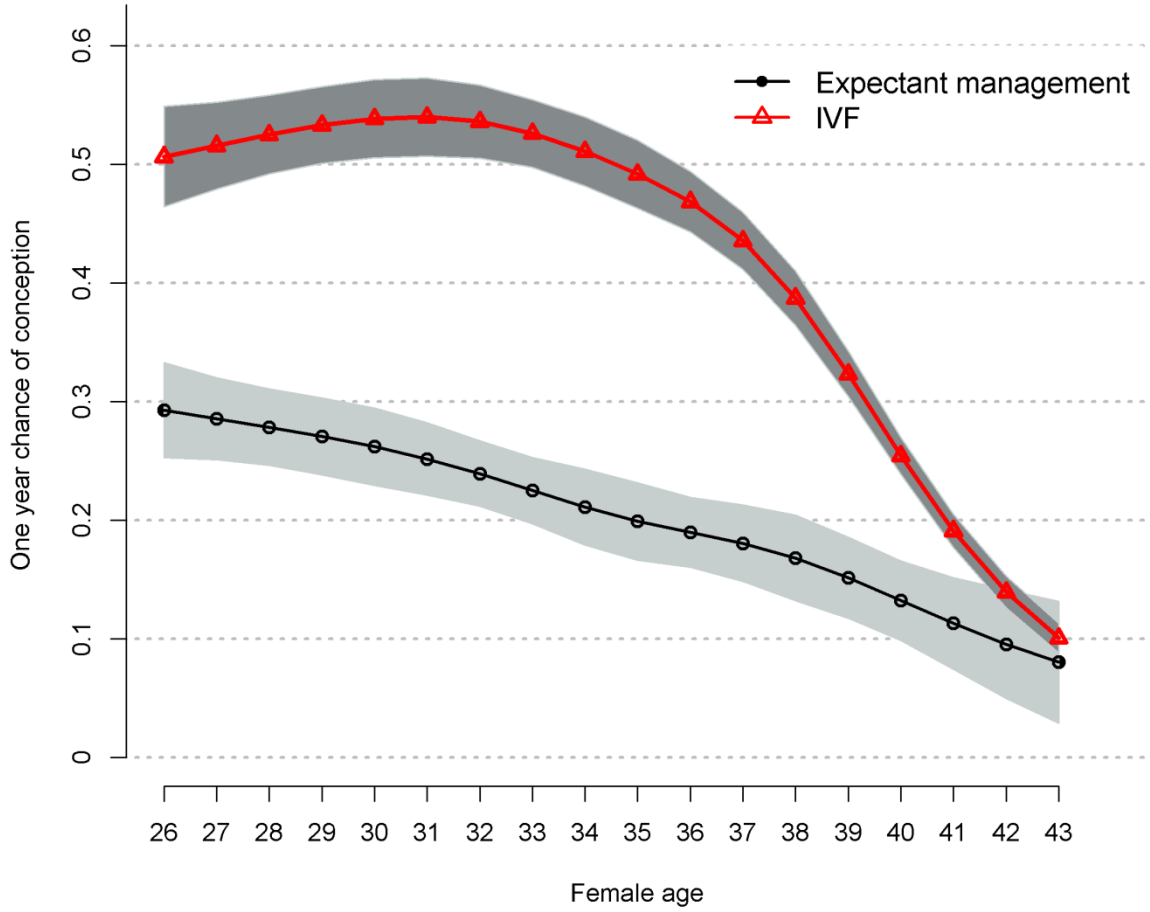


Figure 4. Association between duration of subfertility and the one year chance of conception receiving IVF or pursuing expectant management for a primary subfertile couple of which the woman is 35 years old. Grey bands are 95% confidence limits