Is intrauterine insemination with ovarian stimulation effective in couples with unexplained subfertility?

Journal:	*Human Reproduction*
Manuscript ID	HUMREP-18-0268.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	14-Aug-2018
Complete List of Authors:	van Eekelen, Rik; Academic Medical Center, Centre for Reproductive Medicine; University Medical Center Utrecht, Department of Biostatistics and Research Support, Julius Center van Geloven, Nan; Leiden University Medical Center, Medical Statistics, Department of Biomedical Sciences van Wely, Madelon; Academic Medical Center, Centre for Reproductive Medicine McLernon, David; University of Aberdeen, Medical Statistics, Institute of Applied Health Sciences Mol, F.; Academic Medical Center, Centre for Reproductive Medicine Custers, Inge M.; Academic Medical Center, Centre for Reproductive Medicine Steures, Pieternel; Jeroen Bosch Ziekenhuis, Department of Obstetrics and Gynaecology Bhattacharya, Siladitya; University of Aberdeen, Institute of Applied Health Sciences Mol, Ben; Monash University, Department of Obstetrics and Gynaecology van der Veen, Fulco; Academic Medical Center, Centre for Reproductive Medicine Eijkemans, Marinus; University Medical Center Utrecht, Department of Biostatistics and Research Support, Julius Center
Keywords:	Intrauterine insemination, Unexplained subfertility, Prospective cohort, Time-varying treatment, Sequential Cox model
Subject Section:	Infertility

Is intrauterine insemination with ovarian stimulation effective in couples with unexplained subfertility?

Running title: Effectiveness of IUI with ovarian stimulation

van Eekelen R[1,2,*], van Geloven N[3], van Wely M[1], McLernon DJ[4], Mol F[1], Custers IM[1], Steures P[5], Bhattacharya S[6], Mol BW[7], van der Veen F[1], Eijkemans MJ[2]

[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 Sciences, Leiden University Medical Centre, Einthovenweg 20, 2333 ZC Leiden, the Netherlands [4]Medical Statistics Team, Institute of Applied Health Sciences, University of Aberdeen, AB24 3FX Aberdeen, United Kingdom [5]Department of Obstetrics and Gynaecology, Jeroen Bosch Ziekenhuis, Henri Dunantstraat 1, 5223 GZ Den Bosch, the Netherlands [6]Cardiff University School of Medicine, Heath Park Cardiff CF14 4XN, United Kingdom [7]Department of Obstetrics and Gynaecology, Monash University, Scenic Blvd, VIC 3800 Clayton, Australia

*Correspondence address. E-mail: r.vaneekelen@amc.uva.nl

Extended abstract

Study question: Does starting intrauterine insemination with ovarian stimulation (IUI-OS) within one and a half years after completion of the fertility workup increase ongoing pregnancy rates compared to expectant management in couples with unexplained subfertility?

Summary answer: IUI-OS is associated with higher chances of ongoing pregnancy compared to expectant management in unexplained subfertile couples, specifically those with poor prognoses of natural conception i.e. <15% over 6 months or <25% over one year.

What is known already: IUI-OS is often the first-line treatment for couples with unexplained subfertility. Two randomised controlled trials compared IUI-OS to expectant management using different thresholds for the prognosis of natural conception as inclusion criteria and found conflicting results. A cohort of couples with unexplained subfertility exposed to expectant management and IUI-OS offers an opportunity to determine the chances of conception after both strategies and to evaluate whether the effect of IUI-OS depends on a couple's prognosis of natural conception.

Study design, size, duration: A prospective cohort study on couples with unexplained or mild male subfertility who could start IUI-OS at any point after completion of the fertility workup, recruited in 7 Dutch centres between January 2002 and February 2004**.** Decisions regarding treatment were subject to local protocols, the judgement of the clinician and the wishes of the couple. Couples with bilateral tubal occlusion, anovulation, or a total motile sperm count $<1x10^6$ were excluded. Follow up was censored at the start of IVF, after the last IUI cycle or at last contact and truncated at a maximum of one and a half years after the fertility workup.

Participants/materials, setting, methods: The endpoint was time to conception leading to an ongoing pregnancy. We used the sequential Cox approach comparing in each month ongoing pregnancy rates over the next 6 months of couples who started IUI-OS to couples who did not. We calculated the prognosis of natural conception for individual couples, updated this over consecutive failed cycles and evaluated whether prognosis modified the effect of starting IUI-OS. We corrected for known predictors of conception using inverse probability weighting.

Main results and the role of chance: Data from 1896 couples were available. There were 800 couples whom had at least one IUI-OS cycle within one and a half years post fertility workup of whom 142 couples conceived (rate: 0.50 per couple per year, median follow up 4 months). The median period between fertility workup completion and starting IUI-OS was 6.5 months. Out of 1096 untreated couples, 386 conceived naturally (rate: 0.31 per couple per year, median follow up 7 months).
Starting IUI-OS was associated with a higher chance of ongoing pregnancy by a pooled, overall hazard ratio of 1.96 (95%CI: 1.47-2.62) compared to expectant management. The effect of treatment was modified by a couple's prognosis of achieving natural conception (p=0.01), with poorer prognoses or additional failed natural cycles being associated with a stronger effect of treatment. The predicted 6-month ongoing pregnancy rate for a couple with a prognosis of 25% at completion of the fertility workup over the next 6 cycles (approximately 40% over one year) was 25% (95%CI: 21-28%) for expectant management and 24% (95%CI: 9-36%) when starting IUI-OS directly. For a couple with a prognosis of 15% (25% over one year), these predicted rates were 17% (95%CI: 15-19%) for expectant management and 24% (95%CI: 15-32%) for starting IUI-OS.

Limitations, reasons for caution: The effect estimates are based on a prospective cohort followed up for one and a half years after completion of the fertility workup. Although we

balanced the known predictors of conception between treated and untreated couples using inverse probability weighting, observational data may be subject to residual confounding. The results need to be confirmed in external datasets.

Wider implications of the findings: These results explain the discrepancies between previous trials that compared IUI-OS to expectant management, but further studies are required to establish the threshold at which IUI-OS is (cost-)effective.

Study funding/competing interest(s): This study was facilitated by grant 945/12/002 from ZonMW, The Netherlands Organization for Health Research and Development, The Hague, The Netherlands. BWM is supported by a NHMRC Practitioner Fellowship (GNT1082548). BWM reports consultancy for ObsEva, Merck and Guerbet. SB reports acting as Editor-in-Chief of HROpen. The other authors have no conflicts of interest.

Keywords

Intrauterine insemination; unexplained subfertility; prospective cohort; time-varying treatment; sequential Cox model

Introduction

Couples who have been trying to conceive for at least 12 months and whose fertility workup fails to reveal any abnormalities are considered to have unexplained subfertility (Aboulghar *et al.*, 2009; Brandes *et al.*, 2010). IUI is often used as first-line treatment in these couples, especially in combination with ovarian stimulation (OS), since it is less invasive and less costly than IVF (Tjon-Kon-Fat *et al.*, 2015), despite conflicting evidence from randomised controlled trials regarding the effectiveness of IUI-OS. The two trials that compared IUI-OS to expectant management used different thresholds for the prognosis of natural conception as inclusion criteria. In women with an intermediate prognosis to conceive naturally i.e. an estimated probability between 30% and 40% to conceive within 12 months leading to live birth, IUI-OS was not more effective than expectant management (Steures *et al.*, 2006). In women with a poor prognosis i.e. <30% over 12 months, IUI-OS resulted in more live births than expectant management (Farquhar *et al.*, 2018).

The results of these two trials suggest that IUI-OS might be effective in couples with a poor prognosis, whereas it might be ineffective in couples with better prognoses. For example, couples who have been trying to conceive for a longer period of time and where the female partner is older and/or nulliparous might derive greater benefit from treatment (McLernon *et al.*, 2014). However, these two trial results cannot be taken as definite evidence that we should treat all couples with a poor prognosis of natural conception. Other -unmeasured- differences between the two trials could also explain the different results. The hypothesis that the prognostic profile of a couple determines the benefit of IUI-OS thus needs to be addressed directly in a single population that is heterogeneous in terms of their prognosis of natural conception. Knowledge on who is more likely to conceive with IUI-OS is critical in informing clinical decision making and avoids unnecessary treatment in some while ensuring early and appropriate access to active treatment in others.

The aim of this study was to determine the chances of conception after expectant management or starting IUI-OS in a cohort of unexplained subfertile couples that included couples who followed both strategies.

Materials and Methods

Patient selection

Couples were selected from a prospective cohort recruited across 38 hospitals in The Netherlands between January 2002 and February 2004, the detailed protocol for which has been described elsewhere (van der Steeg *et al.*, 2007). The purpose of recruiting this cohort was to validate the Hunault model to predict the chances of natural conception leading to live birth (Hunault *et al.*, 2004). In 7 of these 38 centres, data on IUI cycles was also recorded and used to validate a prediction model for chances of ongoing pregnancy per IUI cycle (Steures *et al.*, 2004; Custers *et al.*, 2007). IVF pregnancy outcome data were not routinely collected but starting dates of IVF treatment were known. Couples from the 7 centres that recorded IUI data were included in the current study.

We defined subfertility as couples trying to conceive for at least 12 months (Habbema *et al.*, 2004; Gnoth *et al.*, 2005). Selected subfertile couples had regular menstrual cycles (cycle length between 23 and 35 days), at least one patent fallopian tube if hysterosalpingography, laparoscopy or transvaginal hydrolaparoscopy was performed and a total motile sperm count > 1 x 10^6.

Women were evaluated for tubal patency according to the Dutch national guidelines and protocols from 2002 to 2004 recommending the chlamydia antibody test (CAT) as the first-line test (NVOG, 2004). In women who were CAT negative, invasive diagnostic testing was usually refrained from, thus limiting tubal patency testing to the CAT (Broeze *et al.*, 2011). Women who

tested positive on the CAT, or if their history indicated a high risk of tubal pathology, were subsequently tested with hysterosalpingography, transvaginal hydrolaparoscopy or laparoscopy. Decisions regarding treatment - IUI, IVF or expectant management - were made according to local protocols and subject to the judgement of the clinician and the wishes of the couple. Expectant management was defined as no intervention or monitoring aside from the advice to have intercourse. Eighty-three percent of couples who received IUI used stimulation in at least one cycle and treatment was considered IUI-OS. The IUI protocols have been described in more detail elsewhere (Custers *et al.*, 2007).

Follow up and outcome definitions

For the follow up of selected couples, we distinguished between time spent pursuing expectant management and time spent receiving IUI-OS cycles. The start of the IUI period was defined as the first day of menstruation before the first IUI cycle. The end of the IUI period was defined as the first day of menstruation before the last IUI cycle. With this definition, and because natural conceptions after unsuccessful IUI cycles were not recorded, all pregnancies in the IUI period resulted from IUI. The start of expectant management coincided with the completion of the fertility workup and ended at the last date of contact, first day of last menstruation before starting IUI or IVF or, in case they conceived naturally, the first day of the last menstruation before conceiving. We visualised the transition from expectant management to IUI-OS in the cohort by counting the number of couples in both groups over follow up.

The endpoint was ongoing pregnancy, defined as the presence of foetal cardiac activity at transvaginal sonography at a gestational age of at least 12 weeks (van der Steeg *et al.*, 2007). Couples who miscarried before 12 weeks were not censored since they could still achieve ongoing pregnancy in subsequent cycles after their miscarriage. If no ongoing pregnancy occurred, we censored follow up at the end of expectant management or, if treated, at the end of the IUI period.

Cumulative pregnancy rates over multiple IUI cycles

We opted for the Sequential Cox approach to be able to compare treated and untreated couples over multiple cycles after starting treatment, not only directly after completion of the fertility workup but also if they started later (Gran *et al.*, 2010).

In this approach, we derived multiple datasets from the cohort in which couples started IUI-OS at approximately the same point in time and compared them to couples undergoing expectant management at that time, mimicking hypothetical randomised controlled trials (Gran *et al.*, 2010). At completion of the fertility workup and each consecutive month thereafter, i.e. the landmark time points, we constructed such a new mimicked trial from our data in which we included all couples who remained in the cohort i.e. couples who had not conceived and who were not lost to follow up before that point. In these 'trial' sets spanning 6 months, we considered couples as treated if they started IUI-OS early i.e. within one month after the landmark time point. Couples who started IUI-OS within the 6 month window of a trial, but later than one month after the landmark time point, were 'artificially censored' at the time of starting IUI-OS to retain a treatment group that all started at approximately the same time. This way, couples were not included in a single group throughout the study. Instead, couples who started IUI-OS were analysed as controls (under expectant management) in the trials preceding the month in which they started IUI-OS, at which point they were analysed over cumulative treatment cycles as part of the treated (IUI-OS) group in the mimicked trial that started that month.

The maximum follow up period of 18 months after the fertility workup was chosen because of the small numbers of couples starting treatment thereafter. Thus, we derived trial datasets from landmark time point 0 i.e. the fertility workup until the final landmark time point at 12 months after the workup, sliding forward in intervals of one month, resulting in a total of 13 mimicked trials.

193 *Adjusting for patient characteristics that differed between treated and untreated couples*

194 To balance treated, untreated, artificially censored and uncensored couples in important
195 predictors of conception, we applied iterative inverse probability weighting (Austin, 2011; van der
196 Wal, 2011; Austin and Stuart, 2015). Details on how we derived the weights to adjust for these
197 differences are given in the **Supplementary Material**. The patient characteristics we chose to
198 balance were: female age, duration of subfertility, primary or secondary subfertility, total motile
199 sperm count, referral status, presence of one-sided tubal pathology and fertility clinic (Hunault *et*
200 *al.*, 2004; van Eekelen *et al.*, 2017a). We calculated the mean weight, which is ideally around 1 to
201 avoid inflating the effective sample size (Cole and Hernan, 2008).

202

203 We assessed the degree of balance in patient characteristics before and after weighting using the
204 standardized mean difference between the treated and untreated group. A lower standardized
205 mean difference between groups represents better balance and a value below 0.10 generally
206 indicates no important difference (Austin, 2011; Austin and Stuart, 2015).

207

208 *Statistical analysis*

209 We analysed the weighted datasets using a pooled Cox proportional hazards model with IUI-OS
210 or expectant management as a treatment covariate. We calculated an overall hazard ratio by
211 stratifying on all 13 mimicked trials. We used a robust sandwich variance estimator to adjust
212 precision measures since couples can be included in multiple trials (Wei *et al.*, 1989). Using the
213 pooled Cox model, we predicted the probability of conception over 6 months when couples start
214 IUI-OS immediately after completion of the fertility workup or when they remain on expectant
215 management.

216

217 *Modification of the estimated effect of IUI-OS by the prognosis of natural conception*

To address whether the effect of starting IUI-OS depends on the decreasing prognosis of natural conception of the individual couple, we added a treatment-by-prognosis interaction term to the model. We calculated a time-updated prognosis of natural conception over the next 6 cycles at the start of each mimicked trial dataset by using the dynamic prediction model that comprises female age, duration of subfertility, primary or secondary subfertility, percentage of progressive motile sperm and referral by a general practitioner or specialist (van Eekelen *et al.*, 2017a). The prognosis for a couple is thus not one fixed value throughout the study, but decreases after consecutive failed natural cycles. We included the complementary log-log of this updated prognosis as a main effect, the main effect for treatment and the treatment-by-prognosis interaction effect in the pooled Cox model. The weighting procedure was adjusted slightly for this analysis (VanderWeele, 2009) because the difference in prognosis between groups was adjusted for by adding it to the model as a main effect (see also **Supplementary Material)**.

For three hypothetical couples, we visually depicted the relationship between their worsening prognoses and the accompanying 6-month cumulative predicted probability of conception following expectant management or starting IUI-OS, shown as treatment is initiated later. The first example is a couple referred by their general practitioner, where the female partner is nulliparous and 32 years old, the couple has 1 year of subfertility at the time of completion of the fertility workup and the semen analysis showed 37% progressively motile sperm. In this case, the estimated prognosis of natural conception over the next 6 cycles is 25%. A second couple with the same characteristics except for a 2 year duration of subfertility at the completion of the fertility workup has a prognosis of 20% while a third couple with the same characteristics but for a 3.5 year duration of subfertility has a prognosis of 15%. At the time of the completion of their fertility workup, these couples have prognoses of 25%, 20% and 15% respectively over 6 cycles, which translates to approximately 40%, 32% and 25% respectively over 13 cycles i.e. one year (van Eekelen *et al.*, 2017a).

If these three hypothetical couples should fail to conceive naturally over the course of one year after completion of the fertility workup i.e. if they would 'enter' the latest mimicked trial, their prognoses over the next 6 cycles decrease to 13%, 10% and 7% respectively.

Estimated cumulative probabilities of ongoing pregnancy from this model are derived from the separate mimicked trials that all have different observed conception rates, thus predictions may fluctuate. We considered an absolute difference of more than 5% between point estimates of the cumulative ongoing pregnancy rates, estimated at the completion of the fertility workup, to signify a benefit of IUI-OS.

In addition to the impact of prognosis or a failure to conceive in consecutive natural cycles on the effect of treatment, we modelled if the effect of IUI-OS depends on the time of initiation of treatment by adding an additional interaction between treatment and landmark time point to the pooled Cox model already including treatment, prognosis and the treatment-by-prognosis interaction. We also added a three-way interaction between treatment, prognosis and landmark time point to the previous model to see if the effect modification of prognosis on IUI-OS changed over mimicked trials i.e. when starting treatment later.

As a sensitivity analysis, we tested whether couples with mild male subfertility had more or less benefit from IUI-OS compared to couples that did not have mild male subfertility. We classified couples as having mild male subfertility when they had a total motile count between 1 and $10x10^6$, then tested the hypothesis by fitting a Cox model including treatment, the mild male subfertility classification and their interaction.

We used Akaike's Information Criterion (AIC) and Wald tests for the interaction terms to determine whether including the interactions resulted in a better fit of the model to the data (Akaike, 1974).

Missing data

Missing data were accounted for using multiple imputation in a previous study, creating ten imputation sets (van Eekelen *et al.*, 2017a). In total, only 1.3% of patient characteristic data used for this study were missing and we chose to use one randomly selected imputation set for our analyses.

All statistical analyses were performed using R version 3.3.2 [2222](R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/) using the *survival*, *dynpred* and *CreateTableOne* packages.

Results

From the 7860 couples included in the initial cohort, we selected 1896 couples for analysis (Figure 1).

Of these, 800 couples had at least one IUI cycle within one and a half years after the workup for a total of 3119 cycles (**Table I**) following which 142 couples conceived leading to ongoing pregnancy (rate: 0.50 per couple per year over a median follow up for IUI of 4 months). Out of 1096 untreated couples followed up for one and a half years after the fertility workup, 386 conceived naturally leading to ongoing pregnancy (rate: 0.31 per couple per year over a median follow up of 7 months). Among couples who remained on expectant management, there were 5 multiple pregnancies, 68 women miscarried and 7 had an ectopic pregnancy. Among couples treated with IUI-OS, there were 16 multiple pregnancies, 35 women miscarried and 7 women had an ectopic pregnancy.

Out of 800 couples who underwent IUI-OS, 64 started treatment directly after completion of the fertility workup and 736 had a prior period of expectant management after completion of their

fertility workup. The median period between completion of the fertility workup and starting IUI-OS was 6.5 months. In 57% of IUI cycles, recombinant gonadotrophins were used, in 24% no medication was used, in 9% clomiphene citrate was used and in 7% urinary gonadotrophins were used. In 3% of cycles, another type of drug was used or data were missing. Two hundred and sixty eight couples (14%) received IVF as their first treatment, with a median period of expectant management of one year between completion of the fertility workup and the start of IVF.

We depicted the number of couples currently in the expectant management or IUI-OS group over time in **Figure 2**. Until ten months after the fertility workup, there was a steady increase in the number of couples who were currently in a treatment pathway, after which this number declined again.

The baseline characteristics for couples who eventually received at least one cycle of IUI-OS within one and a half years after the fertility workup or who remained untreated are summarized in **Table I**. Treated couples more often had primary subfertility (70% versus 60%) and were more often tested for tubal patency using laparoscopy, hysterosalpingography or transvaginal hydrolaparoscopy (55% versus 39%) compared to couples that were not treated. Female age, on average 32.1 years old (SD: 4.4) and median duration of subfertility of 1.6 years (5th-95th percentile: 1.0-4.7), both at completion of the fertility workup, were similar between groups.

In the weighted trial datasets, the standardized mean differences between treated and untreated couples were below 0.10 for all characteristics, indicating that the two groups were well balanced in terms of prognostic factors after weighting. The mean weight used in the pooled dataset was 1.00, indicating that weights are stable and do not artificially inflate sample size.

Effect estimates of IUI-OS

Starting IUI-OS was associated with increased ongoing pregnancy rates compared to expectant management as shown by an estimated hazard ratio of 1.96 (95%CI: 1.47-2.62), pooling all 13 mimicked, weighted trials running over 6 months.

The predicted probability that a couple would conceive over the course of 6 months of expectant management after the fertility workup was 17% (95%CI: 16-19%). If the couple started IUI-OS directly after completion of the fertility workup, their estimated probability of conception was 31% (23-38%) over that same period.

The relative effect of IUI-OS depended on the prognosis of natural conception (p=0.01).

The relationship between prognosis and the estimated treatment effect as time progresses is visualised in **Figure 3**, in which estimated 6-month cumulative probabilities of conception with and without starting IUI-OS are shown for three different example couples with a prognosis to conceive naturally at completion of the fertility workup over the next 6 cycles of 25% (**A**), 20% (**B**) or 15% (**C**), which were updated over time when these couples fail consecutive natural cycles and start treatment later.

The absolute cumulative predicted probability to conceive over 6 months after starting IUI-OS was stable at around 24%, regardless of the prognosis at completion of the workup or the time thereafter when a couple would start, which leads to larger differences between IUI-OS and expectant management for couples at lower prognoses of natural conception and/or for couples after additional failed natural cycles, both on absolute and relative scale.

It follows from **Figure 3** that a couple who has tried to conceive for 3.5 years post fertility work up with a prognosis of 15% over the next 6 months (approximately 25% over one year) has a higher predicted probability when starting treatment directly after completion of the fertility workup compared to expectant management, whereas a couple who tried to conceive for 1 year with a prognosis of 25% (approximately 40% over one year) does not.

Apart from the influence of decreasing prognosis over time, we found no evidence that the relative effect of IUI-OS changed as treatment was started later (p=0.38) nor that the dependency of the relative effect of IUI-OS on prognosis additionally depended on when treatment was started (p=0.66). We found no evidence for a difference in the benefit of IUI-OS between couples with or without mild male subfertility (p=0.75).

Discussion

In couples with unexplained subfertility, starting IUI-OS within one and a half years after completion of the fertility workup was associated with increased ongoing pregnancy rates over 6 months compared to expectant management. The estimated benefit of treatment depended on the prognosis of natural conception, not only expressed at the time of completion of the fertility workup but also after additional failed natural cycles.

Our study has a number of strengths. First, the chosen study design represents the best possible design for an observational study, being less subject to selection bias than a comparison of separate cohorts since unexplained subfertile couples that will eventually be treated or not were both sampled using a prospective approach and followed up thereafter (Aboulghar *et al.*, 2009; McLernon *et al.*, 2014; van Eekelen *et al.*, 2017b). In addition, we were able to adjust for differences in prognostic factors between treated and untreated couples and the relatively large sample size and long follow up meant we were able to study whether the prognosis of natural conception influences the effect of IUI, an approach aimed to improve the management of unexplained subfertile couples and to explain the discrepancies between existing trials.

Second, data collection took place before the current Dutch clinical guideline for fertility recommending the application of the prediction model for natural conception leading to live birth

by Hunault *et* al. to decide on treatment (NVOG, 2010). Because decisions regarding treatment were not clear-cut at the time, this led to considerable variation in observed time points when couples started IUI and ensured that most couples would be eligible for treatment, reducing confounding by indication.

Third, we defined the IUI period as the time between the first and last cycle, even when treatment cycles were not consecutive, which is a realistic measure of the actual time that couples spend in an IUI program. This allowed comparison with expectant management on the same axis representing 'real' calendar time and allows our results to be easily interpreted (Daya, 2005).

Fourth, we focused on the clinical question on what would occur when treatment would be started at a given time point and applied methodology that matched this question. When analysing the data using a regular Cox approach with a time-varying covariate for treatment, the couples currently undergoing IUI at a certain point in time consist of couples who just started and couples who already failed several cycles, such that the resulting estimate would be different from 'starting IUI' and more difficult to interpret (Gran *et al.*, 2010).

Our study also has some limitations. An adequately powered, randomised controlled trial offers the best way of evaluating the effectiveness of IUI-OS compared to expectant management; our observational study following a cohort of couples after both strategies is a less robust design because treatment was not randomly allocated, which means that treated and untreated couples might differ on important factors related to conception. We tried to balance these factors in both groups using (iterative) inverse probability weighting. Nevertheless, our results could still be influenced by unmeasured factors related to conception i.e. residual confounding. However, the advantage of using observational data on couples with a wide range of prognoses of natural conception was that we could directly address the hypothesis regarding the influence of prognosis on the relative effect of IUI-OS. Addressing this hypothesis with an experimental

design would require the unfeasible, inefficient and indirect approach of conducting multiple smaller, separate trials applying different inclusion criteria.

The primary outcome was ongoing pregnancy since following couples for live birth increases logistical efforts that the study budget did not allow for. In addition, following couples for live birth increases the likelihood of loss-to-follow-up. Ongoing pregnancy is generally considered an appropriate and efficient proxy for live birth in clinical research: approximately 95% of ongoing pregnancies lead to live birth (Clarke *et al.*, 2010; Braakhekke *et al.*, 2014).

54% of couples only received a CAT and not a visual test for tubal patency. It is unlikely that this has led to clinically relevant differences in tubal pathology between treated and untreated couples because the decision to initiate IUI-OS was taken in the absence of visual confirmation of tubal patency and thus, any difference is likely due to chance.

We lacked the sample size to accurately estimate interactions between all separate patient characteristics female age, duration of subfertility etc. and IUI-OS to further individualize treatment effects. Instead, we opted for the summary score in terms of an estimated prognosis such that only one additional parameter was required (Dahabreh *et al.*, 2016) and updated this over time using our previously developed dynamic model for natural conception to determine when a couple would benefit from starting treatment (van Eekelen et al., 2017a).

Our findings add context to the apparent discrepancies between the two randomised controlled trials that compared IUI-OS to expectant management (Steures *et al.*, 2006; Farquhar *et al.*, 2018). Steures *et al*. included 253 couples with an intermediate prognosis for natural conception i.e. 30% to 40% as calculated by the Hunault model and found a risk ratio of 0.85 (95%CI: 0.63-1.10) for ongoing pregnancy after IUI-OS compared to expectant management over the course of 6 months. The pragmatic TUI trial by Farquhar *et al*. included 201 couples with a prognosis <30% and found a beneficial effect of IUI-OS compared to expectant management by a risk ratio of 3.41 (95%CI: 1.71-6.79). Couples in the latter study had a much worse prognosis of natural

conception compared to the first; mostly due to an average duration of subfertility of 3.6 years that was almost double the average duration in the trial by Steures *et al*. and in the present study. Our finding that the benefit of IUI-OS is larger in couples with poorer prognoses of natural conception, both on relative and absolute scale, may primarily explain and support the difference in results.

Couples with a prognosis of natural conception of 25% over 6 months (approximately 40% over one year) or higher were not expected to benefit from IUI-OS. For couples with a prognosis around 20% over 6 months (approximately 32% over one year), it was uncertain if we can expect a clinically relevant benefit of IUI-OS. Couples with a prognosis of 15% over 6 months (approximately 25% over one year) or lower were expected to benefit from IUI-OS as the absolute difference in pregnancy rates was more than 5%.

Since IUI-OS is invasive, expensive and associated with potential disadvantages such as ovarian hyperstimulation or multiple pregnancies (Rooney and Domar, 2016), we believe that counselling couples should involve a discussion about their prognosis when following alternative treatment scenarios so that they are able to make an informed choice. These results, in combination with the prognosis of natural conception derived from a prediction model (van Eekelen *et al.*, 2017a), can be used to attain that goal.

Because confidence intervals around the predicted pregnancy rates after IUI-OS for a given prognosis were broad in the current study, the results need to be interpreted with caution and replicated in further research.

Conclusion

Within one and a half years post fertility workup, starting IUI-OS is associated with higher chances of conception in couples with unexplained subfertility and a poor prognosis for natural conception i.e. <15% over 6 months or <25% over one year. Our results explain the

discrepancies between two trials that compared IUI-OS to expectant management using different inclusion criteria for the prognosis of natural conception. These results may be used in counselling couples with unexplained subfertility. Future studies should focus on establishing the threshold for the prognosis of natural conception at which IUI-OS is deemed (cost-)effective.

Acknowledgements

The authors would like to thank Prof. Hein Putter and colleagues from Leiden University Medical Center, the Netherlands and Dr. Marijn Hazelbag and colleagues from the Julius Center, University Medical Center Utrecht, the Netherlands for their insightful comments regarding the methodology used in this paper.

Authors' roles:

NvG, PS, FvdV and MJE conceived the study. PS and IMC collected and cleaned the data. RvE, MJE and NvG designed the statistical analysis plan. RvE analysed the data. RvE and NvG drafted the manuscript. All authors contributed critical revision to the paper and approved the final manuscript.

Funding

This study was facilitated by grant 945/12/002 from ZonMW, The Netherlands Organization for Health Research and Development, The Hague, The Netherlands. BWM is supported by a NHMRC Practitioner Fellowship (GNT1082548).

Conflicts of Interest

BWM reports consultancy for ObsEva, Merck and Guerbet. SB reports acting as Editor-in-Chief of HROpen. The other authors have no conflicts of interest.

References

Aboulghar M, Baird D, Collin J, Evers J, Fauser B, Lambalk C and Other A. Intrauterine insemination. *Hum Reprod Update* 2009;**15**:265-277.

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

Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* 2011;**46**:399-424.

Austin PC and Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med* 2015;**34**:3661-3679.

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.

Brandes M, Hamilton CJ, de Bruin JP, Nelen WL and Kremer JA. The relative contribution of IVF to the total ongoing pregnancy rate in a subfertile cohort. *Hum Reprod* 2010;**25**:118-126.

Broeze KA, Opmeer BC, Coppus SF, Van Geloven N, Alves MF, Anestad G, Bhattacharya S, Allan J, Guerra-Infante MF, Den Hartog JE*, et al.* Chlamydia antibody testing and diagnosing tubal pathology in subfertile women: an individual patient data meta-analysis. *Hum Reprod Update* 2011;**17**:301-310.

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

Cole SR and Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008;**168**:656-664.

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

Dahabreh IJ, Hayward R and Kent DM. Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. *Int J Epidemiol* 2016;**45**:2184-2193.

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

519 Farquhar CM, Liu E, Armstrong S, Arroll N, Lensen S and Brown J. Intrauterine insemination with
520 ovarian stimulation versus expectant management for unexplained infertility (TUI): a pragmatic,
521 open-label, randomised, controlled, two-centre trial. *Lancet* 2018;**391**:441-450.

522

523 Gnoth C, Godehardt E, Frank-Herrmann P, Friol K, Tigges J and Freundl G. Definition and
524 prevalence of subfertility and infertility. *Hum Reprod* 2005;**20**:1144-1147.

525

526 Gran JM, Roysland K, Wolbers M, Didelez V, Sterne JA, Ledergerber B, Furrer H, von Wyl V and
527 Aalen OO. A sequential Cox approach for estimating the causal effect of treatment in the presence
528 of time-dependent confounding applied to data from the Swiss HIV Cohort Study. *Stat Med*
529 2010;**29**:2757-2768.

530

531 Habbema JD, Collins J, Leridon H, Evers JL, Lunenfeld B and te Velde ER. Towards less
532 confusing terminology in reproductive medicine: a proposal. *Hum Reprod* 2004;**19**:1497-1501.

533

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

537

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

541

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

543

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

546

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

549

550 Rooney KL and Domar AD. The impact of stress on fertility treatment. *Curr Opin Obstet Gynecol*
551 2016;**28**:198-201.

552

553 Steures P, van der Steeg JW, Hompes PG, Habbema JD, Eijkemans MJ, Broekmans FJ,
554 Verhoeve HR, Bossuyt PM, van der Veen F and Mol BW. Intrauterine insemination with controlled
555 ovarian hyperstimulation versus expectant management for couples with unexplained subfertility
556 and an intermediate prognosis: a randomised clinical trial. *Lancet* 2006;**368**:216-221.

557

558 Steures P, van der Steeg JW, Mol BW, Eijkemans MJ, van der Veen F, Habbema JD, Hompes
559 PG, Bossuyt PM, Verhoeve HR, van Kasteren YM*, et al.* Prediction of an ongoing pregnancy after
560 intrauterine insemination. *Fertil Steril* 2004;**82**:45-51.

561

562 Tjon-Kon-Fat RI, Bensdorp AJ, Bossuyt PM, Koks C, Oosterhuis GJ, Hoek A, Hompes P,
563 Broekmans FJ, Verhoeve HR, de Bruin JP*, et al.* Is IVF-served two different ways-more cost-
564 effective than IUI with controlled ovarian hyperstimulation? *Hum Reprod* 2015;**30**:2331-2339.

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

b. van Eekelen R, van Geloven N, van Wely M, McLernon DJ, Eijkemans MJ, Repping S, Steyerberg EW, Mol BW, Bhattacharya S and 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 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.

van der Wal W (2011). Causal modelling in epidemiological practice [PhD thesis, 2011]. Chapter 8: Using iterative probability weighting to improve causal effect estimates. University of Amsterdam, Amsterdam, the Netherlands. Available from: http://bit.ly/2kWMrRt.

VanderWeele TJ. On the distinction between interaction and effect modification. *Epidemiology* 2009;**20**:863-871.

Wei L, Lin D and Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* 1989;**84**:1065-1073.

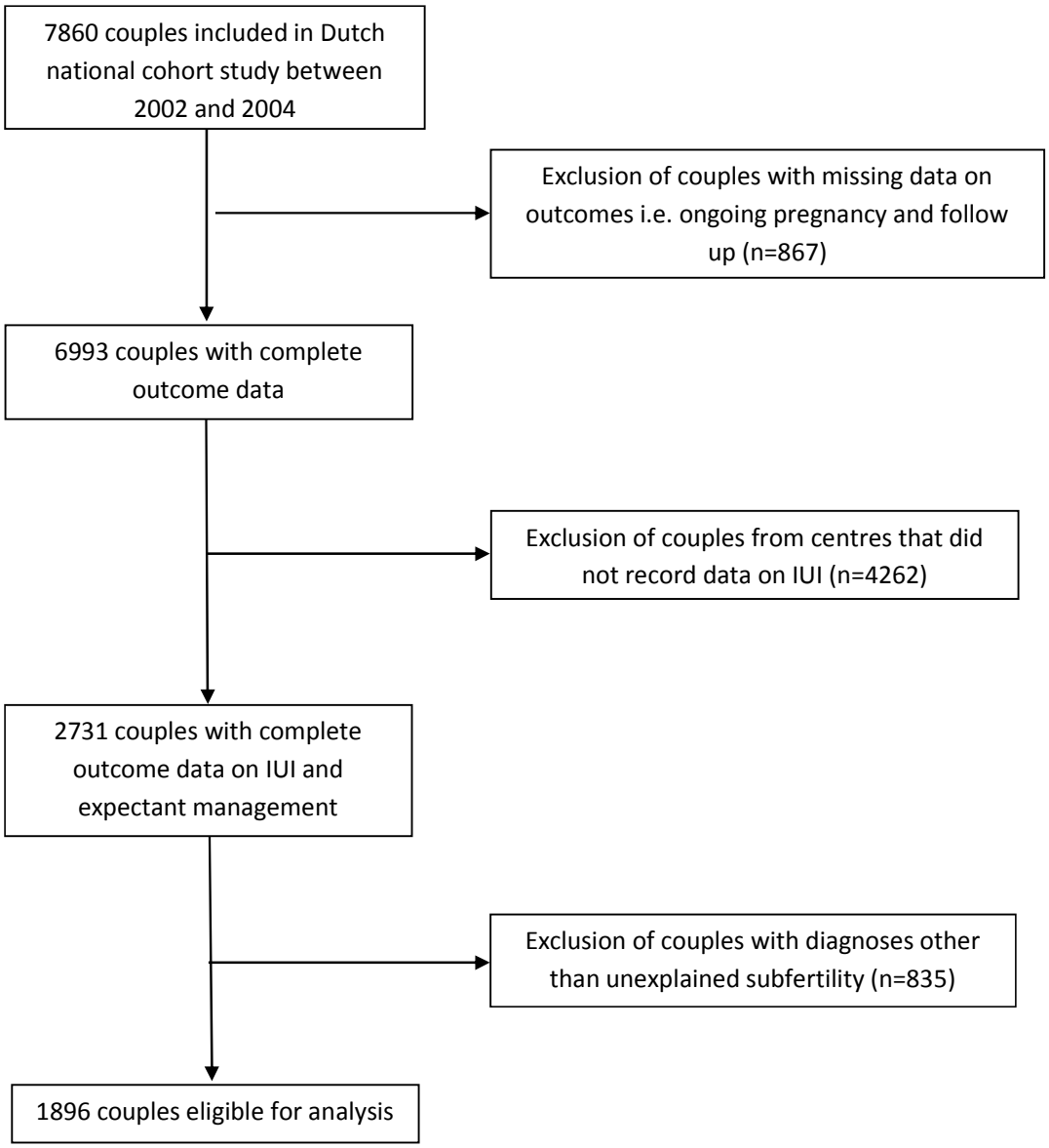
Figure legends

Figure 1 Flow chart of couples from the cohort who were included in the analysis.

Figure 2 Number of couples currently on treatment or not as time progresses, depicted in the original dataset (n=1896) before the sequential Cox procedure.

Figure 3 The association between the predicted prognosis of natural conception and the estimated benefit of starting IUI-OS. This is shown as cumulative probabilities over 6 months (y-axis) when starting IUI-OS or not at different time points after completion of the fertility workup (x-axis) for three example couples that have three different prognoses at the workup completion: 25% (panel A), 20% (panel B) or 15% (panel C). The prognosis was calculated over 6 cycles and updated after additional failed natural cycles. Grey bands represent 95% confidence limits.

Figure 1. Flowchart



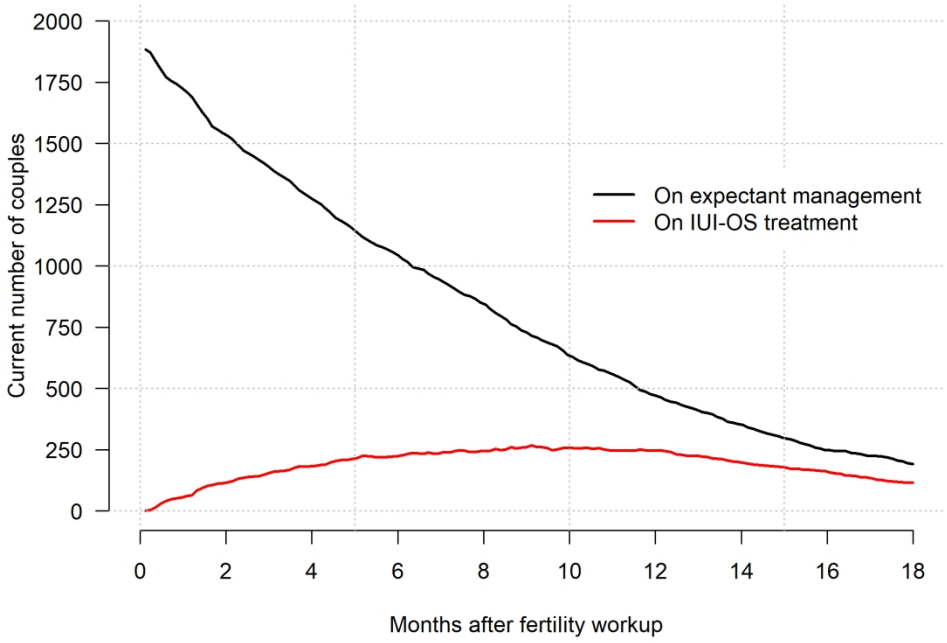


Figure 2. Number of couples currently on treatment or not as time progresses, depicted from the original dataset (n=1896) before the sequential Cox procedure

190x147mm (300 x 300 DPI)

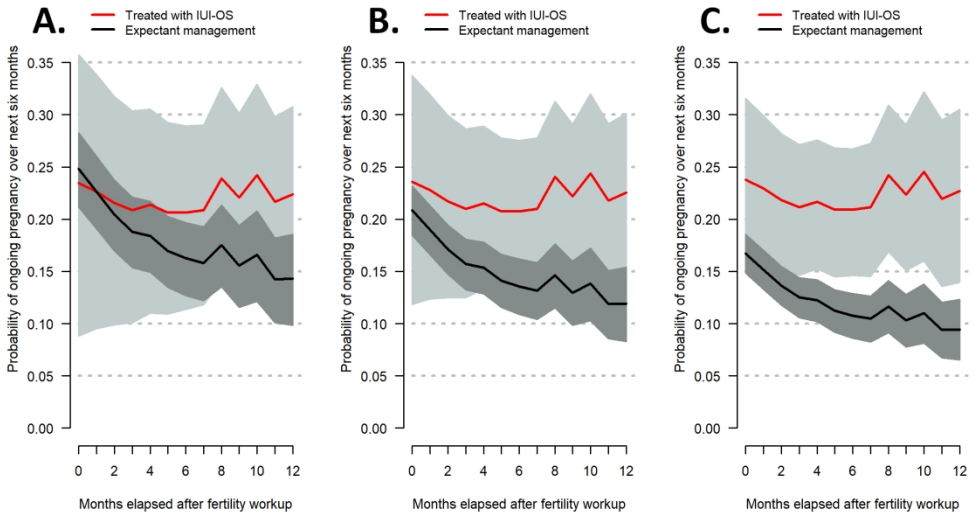


Figure 3. Association between the predicted prognosis of natural conception and the estimated benefit of IUI-OS, shown as cumulative probabilities over 6 months when starting IUI-OS or not for three different couples' prognoses at completion of the fertility workup: 25% (A), 20% (B) or 15% (C). The prognosis was calculated over 6 cycles and updated after additional failed natural cycles.

634x340mm (96 x 96 DPI)

Table I. Baseline characteristics of patients in the original dataset at completion of the fertility workup (n=1896) and the total number of IUI cycles they underwent. Data are mean or median (5th-95th percentile) or n (%).

	Couples who remained on expectant management (n=1096)	Couples who received at least one IUI cycle within one and a half years (n=800)
Female age at workup (years)	32.2 (24.0-39.4)	32.0 (25.3-38.6)
Duration of subfertility at workup (years, median)	1.6 (1.0-4.9)	1.6 (1.0-4.5)
Primary subfertility (versus secondary)	659 (60%)	558 (70%)
Total motile sperm count (in millions)	72 (2-240)	62 (3-205)
Tubal pathology		
Only CAT performed	667 (61%)	363 (45%)
Invasive testing[a] conducted:		
Both fallopian tubes functional	359 (33%)	380 (48%)
One-sided tubal occlusion	70 (6%)	57 (7%)
Ob/Gyn referral (versus GP referral)	125 (11%)	67 (8%)
Number of IUI cycles		
Cycle 1	-	800 (100%)
Cycle 2	-	673 (84%)
Cycle 3	-	561 (70%)
Cycle 4	-	431 (54%)
Cycle 5	-	355 (44%)
Cycle 6	-	272 (34%)
Cycles 7-13	-	27 (3%)

CAT, chlamydia antibody test; Ob/Gyn, gynaecologist; GP, general practitioner

[a]hysterosalpingography, transvaginal hydrolaparoscopy or laparoscopy

Supplementary material. Details regarding the weighing procedure

In our iterative inverse probability weighting (IIPW) algorithm per trial, we started the iterative process by first calculating the proportion of treated and untreated couples. We then divided this by the predicted chance for individual couples of being treated or untreated based on their characteristics, estimated by a logistic regression model in every trial dataset (Cole and Hernan, 2008; Austin and Stuart, 2015). The result from this division, a weight for individual couples, gives an indication how likely this couple was to have their observed treatment status and if they should have an increased or decreased influence on the analysis relative to other couples. After obtaining the weights in the first iteration, we applied the resulting weights on the trial dataset in the second iteration, derived new weights in the same manner on the weighted dataset, multiplied the old and the new weights and repeated the process until all measured predictors were balanced between treated and untreated couples (Austin, 2011; van der Wal, 2011). In addition, a second criterion was that the set of weights should balance measured predictors between artificially censored and uncensored couples as well, for which we used a separate logistic regression model predicting censoring, which modified the same set of weights.

We used a maximum of 2000 iterations per mimicked trial dataset. Weights were truncated at the 1st and 99th percentile between iterations to remove extreme values and multiplied with the inverse of the mean weight to center the weights around 1 (van der Wal, 2011). The criterion for stopping the IIPW process was when the variance of the natural log of newly derived weights, for both treatment initiators and censoring, were below 1×10^{-7}, indicating that all new weights are 1 representing no further confounding present in the weighted dataset (van der Wal, 2011), or when the maximum number of iterations was reached. The algorithm is summarized in the **Explanatory Figure**.

We calculated the mean weight of the final set, which is ideally around 1 to avoid inflating the effective sample size (Cole and Hernan, 2008).

We applied the above algorithm in every mimicked trial using the following predictive patient characteristics: female age, duration of subfertility, primary or secondary subfertility, total motile sperm count, referral status, one-sided tubal pathology and fertility clinic. If weights did not converge in a landmark, if they were highly unstable (i.e. no treated couples left in the trial dataset) or if there was residual imbalance in terms of a standardized mean difference above 0.1, weights were re-estimated for that landmark without the location variable.

For the secondary analysis investigating whether the prognosis of natural conception influences the estimated relative effect of IUI-OS, we calculated a different set of weights using female age, duration of subfertility, primary or secondary subfertility, total motile sperm count and referral status in the first (numerator) and divided the resulting probability this time by the probability from a second (denominator) model containing the same variables plus one-sided tubal pathology and fertility clinic. This was because balance in tubal pathology and clinic between groups was desired, but not for the other mentioned variables, since we already accounted for these in the outcome (Cox) model via the summary measure prognosis plus the interaction between prognosis and IUI-OS (vanderWeele, 2009).

Explanatory Figure