

Does the effectiveness of IUI in couples with unexplained subfertility depend on their prognosis of natural conception? A replication of the H2Oil study

R. van Eekelen ^{1,*}, K. Rosielle², N. van Welie², K. Dreyer², M. van Wely¹, B.W. Mol³, M.J. Eijkemans⁴, V. Mijatovic², and N. van Geloven⁵

¹Amsterdam UMC, Academic Medical Centre, Centre for Reproductive Medicine, Amsterdam, the Netherlands ²Department of Reproductive Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands ³Department of Obstetrics and Gynaecology, Monash University, Clayton, VIC, Australia ⁴Department of Biostatistics and Research Support, Julius Centre, University Medical Centre Utrecht, Utrecht, the Netherlands ⁵Medical Statistics, Department of Biomedical Data Sciences, Leiden University Medical Centre, Leiden, the Netherlands

*Correspondence address. Amsterdam UMC, Academic Medical Centre, Centre for Reproductive Medicine, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands. E-mail: r.vaneeekelen@amsterdamumc.nl  <https://orcid.org/0000-0002-9045-3630>

Submitted on April 7, 2020; resubmitted on August 4, 2020; editorial decision on August 10, 2020

STUDY QUESTION: Can we replicate the finding that the benefit of IUI-ovarian stimulation (IUI-OS) compared to expectant management for couples with unexplained subfertility depends on the prognosis of natural conception?

SUMMARY ANSWER: The estimated benefit of IUI-OS did not depend on the prognosis of natural conception but did depend on when treatment was started after diagnosis, with starting IUI-OS later yielding a larger absolute and relative benefit of treatment.

WHAT IS KNOWN ALREADY: IUI-OS is often the first-line treatment for couples with unexplained subfertility. Two randomized controlled trials (RCTs) compared IUI-OS to expectant management using different thresholds for the prognosis of natural conception as inclusion criteria and found different results. In a previous study (a Dutch national cohort), it was found that the benefit of IUI-OS compared to expectant management seemed dependent on the prognosis of natural conception, but this finding warrants replication.

STUDY DESIGN, SIZE, DURATION: We conducted a secondary analysis of the H2Oil study ($n = 1119$), a multicentre RCT that evaluated the effect of oil-based contrast versus water-based contrast during hysterosalpingography (HSG). Couples were randomized before HSG and followed up for 3–5 years. We selected couples with unexplained subfertility who received HSG and had follow-up or pregnancy data available. Follow-up was censored at the start of IVF, after the last IUI cycle or at last contact and was truncated at a maximum of 18 months 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 the 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 975 couples were available. There were 587 couples who received at least one IUI-OS cycle within 18 months after HSG of whom 221 conceived leading to an ongoing pregnancy (rate: 0.74 per couple per year over a median follow-up for IUI of 5 months). The median period between HSG and starting IUI-OS was 4 months. Out of 388 untreated couples, 299 conceived naturally (rate: 0.56 per couple per year over a median follow-up of 4 months). After creating our mimicked trial datasets, starting IUI-OS was associated with a higher chance of ongoing pregnancy by a pooled, overall hazard ratio of 1.50 (95% CI: 1.19–1.89) compared to expectant management. We did not find strong evidence that the effect of treatment was modified by a

couple's prognosis of achieving natural conception (Akaike's Information Criterion (AIC) decreased by 1 point). The effect of treatment was dependent on when couples started IUI-OS (AIC decreased by more than 2 points). The patterns of estimated absolute chances over time for couples with increasingly better prognoses were different from the previous study but the finding that starting later yields a larger benefit of treatment was similar. We found IUI-OS increased the absolute chance of pregnancy by at least 5% compared to expectant management. The absolute chance of pregnancy after IUI-OS seems less variable between couples and starting times of treatment than the absolute chance after expectant management.

LIMITATIONS, REASONS FOR CAUTION: This is a secondary analysis, as the H2Oil trial was not designed with this research question in mind. Owing to sample size restrictions, it remained difficult to distinguish between the ranges of prognoses in which true benefit was found.

WIDER IMPLICATIONS OF THE FINDINGS: We replicated the finding that starting IUI-OS later after diagnosis yields a larger absolute and relative benefit of treatment. We did not replicate the dependency of the effect of IUI-OS on the prognosis of natural conception and could not identify clear thresholds for the prognosis of natural conception when IUI-OS was and/or was not effective. Because many of these couples still have good chances of natural conception at the time of diagnosis, we suggest clinicians should advise couples to delay the start of IUI-OS for several months to avoid unnecessary treatment.

STUDY FUNDING/COMPETING INTEREST(S): The H2Oil study (NTR 3270) was an investigator-initiated study that was funded by the two academic institutions (AMC and VUmc) of the Amsterdam UMC. The follow-up study (NTR 6577) was also an investigator-initiated study with funding by Guerbet, France. The funders had no role in study design, collection, analysis and interpretation of the data. B.W.M. is supported by an Investigator grant (GNT1176437) from the Australian National Health and Medical Research Council (NHMRC). K.D. reports receiving travel and speaker fees from Guerbet. B.W.M. reports consultancy for ObsEva, Merck, Merck KGaA, iGenomix and Guerbet. V.M. reports receiving travel- and speaker fees as well as research grants from Guerbet.

Key words: IUI / unexplained subfertility / natural conception / prospective cohort / expectant management / randomized controlled trial / prognosis / time-varying treatment / sequential Cox / Cox proportional hazards model

WHAT DOES THIS MEAN FOR PATIENTS?

IUI can be the first-line treatment for couples who do not conceive within 1 year of trying and in whom no barrier to conception could be found during the fertility workup. These couples can often still conceive naturally (without medical help) if they try for a little while longer and it is unknown whether insemination offers much benefit to begin with. In a previous study, we found that the benefit of insemination over natural conception differs between couples based on their prognosis, which is the chance of a natural pregnancy for their specific situation calculated using female age and previous pregnancies, among other factors. In the current paper, we aimed to replicate this finding in a different population. We did not find the same variation in treatment effect based on a couple's prognosis as in our previous study, but we did find that the later a couple starts with insemination, the higher their treatment benefit is expected to be. This might be a reason for couples to keep trying to conceive naturally for a little while longer before starting insemination treatment.

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). In several countries, IUI used as first-line treatment in these couples, especially in combination with ovarian stimulation (OS), since IUI is less invasive and less costly than IVF (Tjon-Kon-Fat et al., 2015), despite the lack of evidence from randomized controlled trials (RCTs) regarding the effectiveness of IUI-OS (Wang et al., 2019). The two trials that compared IUI-OS to expectant management used different thresholds for the prognosis of natural conception as inclusion criteria (Steures et al., 2006, Farquhar et al., 2018). 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 no more effective than expectant management (Steures et al.,

2006). In women with a poor prognosis, i.e. <30% over 12 months, IUI-OS did result in more live births than expectant management (Farquhar et al., 2018).

In a previous study, we found that the different outcomes of these two trials might be explained by the difference in the prognosis of natural conception (van Eekelen et al., 2019). In a Dutch cohort of 1896 couples, we found that couples with lower prognoses of natural conception had more benefit from IUI-OS in terms of a relative and absolute difference in the chance of conception compared to expectant management. Due to sample size limitations, it was difficult to identify a fixed threshold for prognoses at which point IUI-OS becomes effective. For a prognosis below 25% over 1 year, IUI-OS seemed effective, leading to higher chances of ongoing pregnancy in 6 months compared to expectant management. For higher prognosis than 40% over 1 year, IUI-OS was not effective and led to similar chances of ongoing pregnancy compared to expectant management. Between these thresholds, it was uncertain whether IUI-OS was effective.

Replication of research findings is an essential part of medical research as many findings unfortunately cannot be reproduced in further studies (Ioannidis, 2005). Addressing the same research question with different methodological approaches, such as trial and observational data, provides more evidence that a single result is not a chance finding (Lawlor *et al.*, 2016; Munafò and Smith, 2018). More knowledge on who does and who does not benefit from IUI-OS can guide clinical practice and inform evidence-based shared decision-making on when to start treatment. Because these thresholds hold great importance to patients, they should be based on solid evidence.

The aim of this study was to replicate the previous result (i.e. that the benefit of IUI-OS compared to expectant management for couples with unexplained subfertility depends on the prognosis of natural conception) in an independent data source derived from an RCT on contrast fluid used for hysterosalpingography (HSG). This is a different approach to see if the effectiveness of IUI-OS here also depends on the prognosis of natural conception, as well as to validate the thresholds of 25% and 40%.

Materials and methods

The H2Oil study was a multicentre RCT to compare ongoing pregnancy rates in subfertile women who underwent HSG with oil-based versus water-based contrast (Netherlands Trial Register number, NTR3270). The study was approved by the Institutional Review Board of the Amsterdam University Medical Centre-Academic Medical Centre (reference 2008.362, dated 12 February 2009). The study details and results have been published previously (Dreyer *et al.*, 2017).

The H2Oil follow-up study (NTR 6577) assessed the long-term outcomes until 3–5 years after the H2Oil study and was approved by the Institutional Review Board of the Amsterdam University Medical Centre, location VU University Medical Centre (reference 2017.221, dated 14 June 2017). Study details and results have been published elsewhere (van Rijswijk *et al.*, 2020).

Inclusion and exclusion criteria

Included women were between 18 and 39 years of age, were having ovulatory cycles and had a low perceived risk of tubal pathology based on their medical history. They had tried to conceive unsuccessfully for at least 1 year and had an indication for tubal patency testing. Exclusion criteria were known endocrine disorders and having a partner with severe male subfertility (defined as a total motile sperm count after sperm wash of less than 3 million per millilitre).

Women were randomized for an HSG with oil-based contrast or an HSG with water-based contrast. Data regarding fertility treatments and pregnancies were collected until 3–5 years after randomization.

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. All pregnancies in the IUI period thus resulted from IUI. Follow-up for

expectant management started 14 days before they received HSG 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.

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 (Dreyer *et al.*, 2017). 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 used the same statistical approach as in our previous study (van Eekelen *et al.*, 2019). In short, we used the sequential Cox approach to compare multiple cycles of treated and untreated couples, 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 RCTs (Gran *et al.*, 2010). At completion of the fertility workup and each consecutive month thereafter, named the landmark time points, we constructed such a mimicked trial from our data in which we included all couples who remained in the cohort, i.e. couples who had not conceived, had not started treatment and were not lost to follow-up before that landmark time point. In these ‘trial’ datasets spanning 6 months, we considered couples as treated if they started IUI-OS early, i.e. within 1 month after the landmark time point. Couples who did not start IUI-OS within the first month were used as controls. Couples who started IUI-OS within the 6 months window of a trial, but later than 1 month after the landmark time point, were counted as controls during their untreated period and ‘artificially censored’ at the time of starting IUI-OS. This way, couples were not included in a single group throughout the study. Instead, couples who at some point started IUI-OS were analysed as controls (under expectant management) in the ‘mimicked’ trial datasets preceding the month in which they started IUI-OS. When they started IUI-OS, their following treatment cycles were analysed as part of the treated (IUI-OS) group in the mimicked trial dataset that started that month.

In order to compare results to the previous study, we restricted our data to a maximum of 18 months of follow-up.

Adjusting for patient characteristics that differed between treated and untreated couples

In our data, couples were not randomized to either expectant management or IUI-OS. Thus, patients starting IUI could differ from those who did not in terms of important predictors of conception such as female age or duration of subfertility. In order to achieve groups that are on average similar, we opted for a statistical technique called applied iterative inverse probability weighting (Austin, 2011; van der Wal, 2011; Austin and Stuart, 2015). By reweighting patients’ contribution to the data, these characteristics are balanced. Details on how we

derived the weights to adjust for these differences are given in the [Supplementary Data](#). We chose to balance for the same patient characteristics as in the previous study with the exception of fertility clinic, as that would lead to very unbalanced weights: female age, duration of subfertility, primary or secondary subfertility, total motile sperm count, referral status and the presence of one-sided tubal pathology (Hunault et al., 2004; van Eekelen et al., 2017a). We calculated the mean weight to assess potential inflation of the effective sample size induced by the weighting, which is ideally around 1 (Cole and Hernan, 2008).

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

Statistical analysis

We analysed the weighted mimicked trial datasets using a pooled Cox proportional hazards model with IUI-OS or expectant management as a treatment covariate. We calculated an overall hazard ratio by stratifying on the 13 mimicked trials. We used a robust sandwich variance estimator to adjust precision measures since couples can be included in multiple mimicked trial datasets (Wei et al., 1989).

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 the prognosis and a treatment-by-prognosis interaction term to the model. We calculated a time-updated prognosis of natural conception over the next six cycles at the start of each mimicked trial dataset by using an existing dynamic prediction model that comprises female age, duration of subfertility, primary or secondary subfertility, percentage of progressive motile sperm, referral by a general practitioner or specialist, and the unsuccessful number of menstrual cycles since the fertility workup (van Eekelen et al., 2017a). The prognosis for a couple that we used is thus not one fixed value throughout the study but decreases after consecutive failed natural cycles. We transformed the updated prognosis by taking the complementary log-log of its value such that it is linear on the log-hazard scale used by the Cox model (Prentice and Gloeckler, 1978). 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 because the difference in prognosis between groups was adjusted for by adding it to the model as a main effect (VanderWeele, 2009) (see also the [Supplementary Data](#)).

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, 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 first six 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 six cycles, which translates to approximately 40%, 32% and 25%, respectively over 13 cycles i.e. 1 year (van Eekelen et al., 2017a).

The chances of natural conception for these three hypothetical couples decrease over time based on the number of unsuccessful menstrual cycles between the diagnosis/HSG and the start of a landmark.

Estimated cumulative probabilities of ongoing pregnancy from this model are derived from the separate mimicked trials that all have different observed conception rates, thus although predictions are expected to decrease over time, our estimates may fluctuate. We considered an absolute difference of more than five percentage points between estimates of the cumulative ongoing pregnancy rates, estimated at completion of the fertility workup, to indicate a benefit of IUI-OS.

In addition to modelling the impact of prognosis and consecutive failed natural cycles on the effect of treatment, we assessed 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. If the interaction between prognosis and treatment yielded a better fit, 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.

We used Akaike's Information Criterion (AIC) (at least two points difference) 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

Data were missing on duration of subfertility ($n=3$), referral status ($n=2$), primary or secondary subfertility ($n=1$), the percentage of progressive motile sperm ($n=81$) and total motile sperm count ($n=93$) and was accounted for using single imputation.

All statistical analyses were performed using R version 3.3.2 (R Core Team (2017). <http://www.R-project.org/>) using the *survival*, *dynpred*, *mice* and *CreateTableOne* packages.

Results

Out of 1119 couples included in the H2O1 trial, we selected 975 for analysis after excluding couples with other diagnoses than unexplained subfertility, couples who conceived before HSG and couples with missing outcome data.

Of these 975, 587 couples (60%) received 2386 IUI cycles after HSG, of whom 221 couples conceived after IUI leading to ongoing pregnancy (rate: 0.74 per couple per year over a median follow-up for IUI of 5 months). Out of 388 couples (40%) followed up for 18 months

of expectant management after HSG, 299 conceived naturally leading to ongoing pregnancy (rate: 0.56 per couple per year over a median follow-up of 4 months).

In total, 62 couples out of 587 (11%) who underwent IUI-OS started treatment directly after HSG and the remaining 525 (89%) first had a period of expectant management. The median period between HSG and starting IUI-OS was 4 months. A total of 1723 (72%) IUI cycles used OS. Forty-two couples (4%) received IVF as their first treatment, with a median period of expectant management of 9 months between completion of the fertility workup and the start of IVF.

We depicted the number of couples followed under expectant management or followed under IUI-OS over time in Fig. 1. Until approximately 6 months of follow-up, the number of couples who were currently in an IUI-OS treatment pathway kept increasing, after which this number declined again.

The baseline characteristics for couples who eventually received at least one cycle of IUI-OS within 18 months after HSG or who remained untreated are summarized in Table I. Treated couples more

often had primary subfertility (73% vs 60%) compared to couples that were not treated. Female age, median duration of subfertility, total motile sperm count, one-sided tubal pathology and referral status were similar between groups.

In the 13 weighted mimicked 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 1.50 (95% CI: 1.19–1.89), pooling all 13 mimicked, weighted trial datasets running over 6 months.

The predicted probability that a couple would conceive over the course of 6 months of expectant management after HSG was 29% (95% CI: 25–32%). If the couple started IUI-OS directly after HSG, their estimated probability of conception in the next 6 months was 40% (32–47%).

Judging by a decrease of at least 2 points in terms of AIC, the relative effect of IUI-OS did not depend on the prognosis of natural conception (AIC decreased by 1 point, $P=0.17$). The relative effect of IUI-OS was dependent on how long after the HSG treatment it was started (AIC decreased by 2.8 points, $P=0.08$).

The relations between prognosis, the start of treatment and the estimated treatment effect are visualized in Fig. 2. The figure shows the 6-month cumulative probabilities of conception with and without starting IUI-OS for three different example couples with a prognosis to conceive naturally at completion of the fertility workup over the next year of 40% (Fig. 2, left panel), 32% (Fig. 2, middle panel) or 25% (Fig. 2, right panel), which were updated over time when these couples fail consecutive natural cycles and start treatment later.

In Fig. 2, the absolute chance to conceive over 6 months decreased over time for expectant management, but not for IUI-OS, of which the absolute chance seemed much less variable between couples and timing of treatment start, at around 37%, 34% and 30% for the three

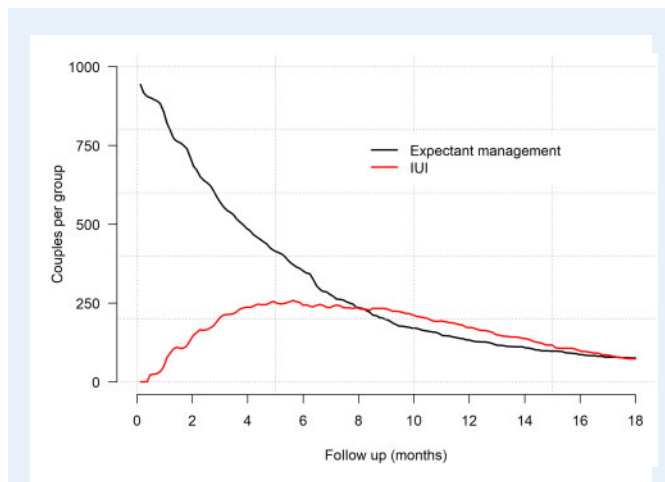


Figure 1. Number of couples who, over follow-up, are currently on expectant management or receiving IUI-OS. IUI-OS, IUI with ovarian stimulation.

Table I Baseline characteristics of 975 patients just before receiving hysterosalpingography.

	Couples who remained on expectant management (n = 388)	Couples who started IUI-OS within 18 months (n = 587)
Female age (years)	32.7 (26.2–38.9)	32.9 (26.1–39.0)
Duration of subfertility (years, median)	1.6 (0.9–4.3)	1.7 (0.9–4.0)
Primary subfertility (vs secondary)	233 (60%)	426 (73%)
Total motile sperm cell count (millions, median)	72 (4–304)	43 (4–294)
One-sided tubal pathology ^a (yes vs no)	16 (4%)	29 (5%)
Referral by specialist (vs referral by general practitioner)	31 (8%)	60 (10%)

Data are mean with the 5th–95th percentile in brackets (unless median is specified) or n (%). OS, ovarian stimulation.

^aAssessed by hysterosalpingography or in addition, a later laparoscopy.

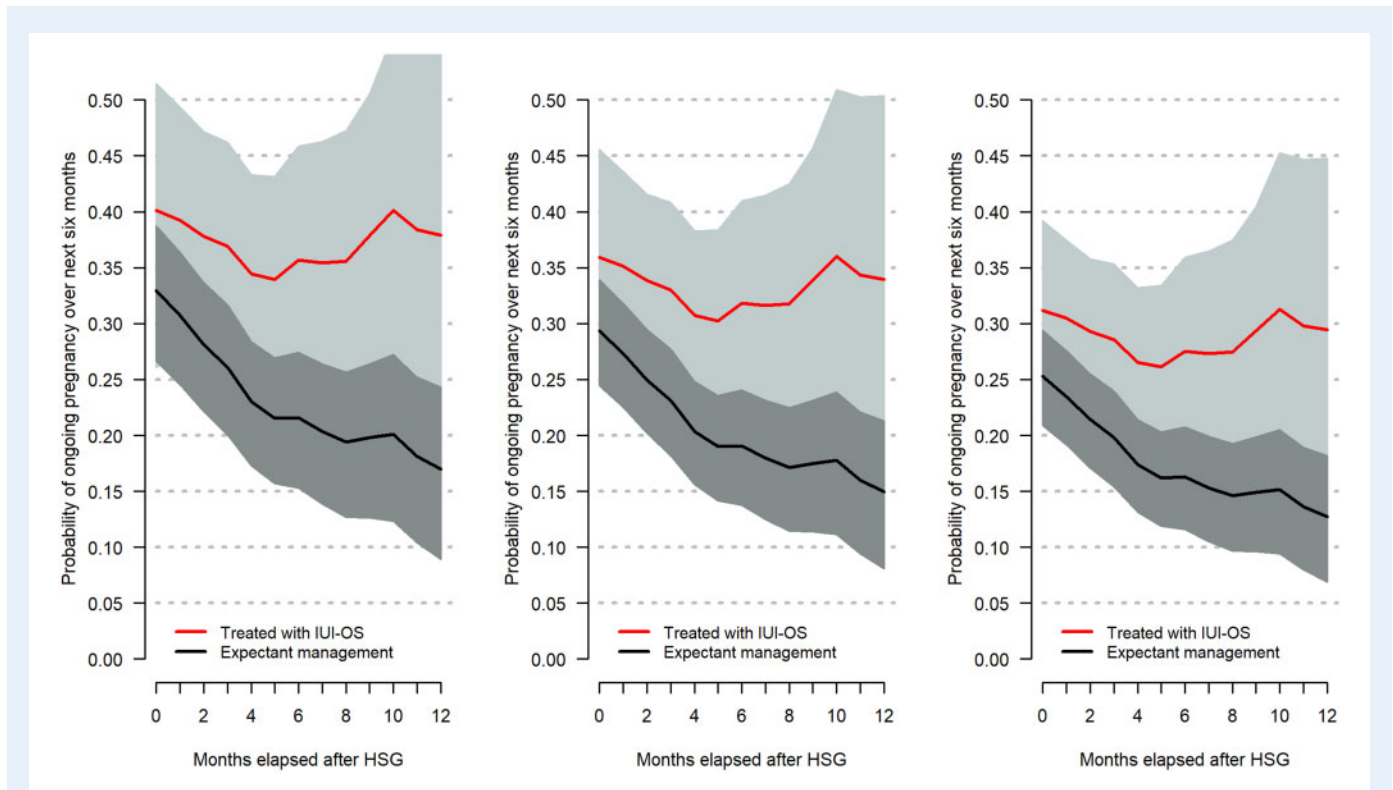


Figure 2. The association between the predicted prognosis of natural conception and the estimated benefit of starting IUI-OS at different time points. This association is shown as cumulative probabilities over 6 months (*y*-axis) when starting IUI-OS, or not, at different time points after completion of the fertility workup (*x*-axis) for three example couples that have three different prognoses at time of hysterosalpingography (HSG): 40% (left), 32% (middle) or 25% (right). The prognosis was calculated over 1 year and updated after additional failed natural cycles. Grey bands represent 95% CIs. Left panel: Couple A is 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. Middle panel: Couple B has the same characteristics as Couple A except for 2 year duration of subfertility at the completion of the fertility workup. Right panel: Couple C has the same characteristics as Couple A but for 3.5 year duration of subfertility.

couples. However, CIs were wide, especially for pregnancy chances after IUI-OS. The decrease in chances for expectant management over time led to a larger treatment benefit as IUI-OS was started later.

It follows from Fig. 2 that the prognosis for a couple does not have a large influence on the expected benefit in terms of the absolute difference between the IUI-OS and expectant management line, as there was always a difference of 5% or more. This was different in the previous study, in which the benefit of IUI-OS was dependent on prognosis. However, the later that treatment was started, the larger the expected benefit of treatment. This was the same finding as in the previous study.

Discussion

We replicated the finding that in couples with unexplained subfertility, starting IUI-OS within 18 months after completion of the fertility workup was associated with increased ongoing pregnancy rates over 6 months compared to expectant management. However, the estimated benefit of treatment did not depend on the prognosis of natural conception but did depend on when treatment was started after

diagnosis. We replicated the finding that starting IUI-OS later yields a larger absolute and relative benefit of treatment.

The main strength of this study was the use of trial data from the H2Oil study with a follow-up of 3–5 years, low loss-to-follow-up and few missing data. For the purpose of triangulation, i.e. the use of multiple approaches to address the same question (Lawlor et al., 2016; Munafò and Smith, 2018), we now have data from two RCTs comparing IUI-OS to expectant management in different patient selection, cohort data and data from an RCT in which couples could receive IUI-OS during follow-up. Workup and treatment protocols differed between these three data sources. Via triangulation, our confidence in two findings has been strengthened: namely that starting IUI-OS later yields a larger absolute and relative benefit of treatment and that the absolute chance of an ongoing pregnancy after IUI-OS is less variable between couples than the chance of natural conception. We did not find significant evidence that the effect of IUI-OS depends on the prognosis for natural conception of a couple.

Weaknesses are that the H2Oil trial was not designed with this secondary question in mind. In addition, the sample size was moderate, which led to wide CIs and perhaps limited the power to show an interaction with prognosis.

We observed that the chances of an ongoing pregnancy after 6 months of expectant management in the present study were much higher than what was found in previous studies, at 29% instead of approximately 18% (Hunault *et al.*, 2004; van der Steeg *et al.*, 2007). This could be due to the fact that in the H2Oil study, all couples received HSG during the diagnostic work-up, which might increase their chances, especially when using oil-based contrast medium (Dreyer *et al.*, 2017, 2019).

The pooled, i.e. overall effect of IUI-OS versus expectant management was less strong in the current study, with a point estimate for the hazard ratio of 1.50 compared to 1.96 that we found previously. The interaction that showed the dependency of the effect of IUI-OS on the prognosis for natural conception was in the same direction as the previous study, with lower prognoses having more benefit of IUI-OS, but this did not reach statistical significance in terms of AIC or *P*-value. It could be that we lacked the power, owing to a less strong main treatment effect and sample size restrictions, to show this dependency as we did in the previous study, which had a larger sample size of 1896 with 800 couples treated with IUI (van Eekelen *et al.*, 2019).

It remains unknown why couples with better prognoses would benefit less from IUI-OS. A possible mechanism is that there is a 'ceiling' to the chance of conception for subfertile couples in terms of a maximum and that some unexplained subfertile couples with good prognoses remain around this ceiling whereas couples with different indications are further below their fertility potential, which is increased by IUI-OS (Moreau *et al.*, 2019).

Not all couples who received IUI-OS conceived or continued IUI-OS over the 6 months follow-up of a 'mimicked' trial dataset. In the previous study, there was no follow-up of natural conception after IUI-OS dropout. We repeated this approach in the current study.

In contrast to the previous study, we found that the effect of IUI-OS depended on when treatment was started independent of the decreasing prognosis of natural conception after failed cycles. This might be due to additional selection over time that is not explained by the dynamic prediction model (van Eekelen *et al.*, 2017a,b, 2019). We found that when accounting for this time effect, the absolute chances after IUI-OS were much more stable over time than chances after expectant management, as the latter clearly decreased over time whereas the former did not. This provides more evidence that the chance after IUI-OS is less dependent on individual factors, i.e. that couples' chances become more similar when receiving IUI-OS. This can be important for counselling couples, as it suggests that further expectant management will not come at a great loss in terms of a decreased chance of pregnancy when receiving IUI-OS later.

Conclusion

We replicated the finding that on average, IUI-OS increases the chance of an ongoing pregnancy compared to expectant management and that when IUI-OS is started later, the expected benefit in terms of the absolute and relative difference with expectant management was larger. We did not replicate the finding that the benefit of IUI-OS depends on the prognosis of natural conception. Couples with unexplained subfertility still have good chances of natural conception at the time of diagnosis, and treatment is thus not always necessary.

Clinicians should counsel couples on the option to prolong expectant management before commencing with IUI-OS.

Supplementary data

Supplementary data are available at *Human Reproduction* Open online.

Authors' roles

N.v.G., B.W.M., V.M., M.v.W. and R.v.E. conceived the study. N.v.W., K.D. and K.R. collected and cleaned the data. R.v.E., N.v.G. and M.J.E. wrote the statistical analysis plan. R.v.E. conducted the statistical analyses. R.v.E., N.v.G., N.v.W. and K.R. drafted the manuscript. All authors contributed critical revision to the paper and approved the final manuscript.

Funding

The H2Oil study (NTR 3270) was an investigator-initiated study that was funded by the two academic institutions (AMC and VUmc) of the Amsterdam UMC. The follow-up study (NTR 6577) was also an investigator-initiated study with funding by Guerbet, France. The funders had no role in study design, collection, analysis and interpretation of the data. B.W.M. is supported by an Investigator grant (GNT1176437) from the Australian National Health and Medical Research Council (NHMRC).

Conflict of interest

K.D. reports receiving travel and speaker fees from Guerbet. B.W.M. reports consultancy for ObsEva, Merck, Merck KGaA, iGenomix and Guerbet. V.M. reports receiving travel- and speaker fees as well as research grants from Guerbet.

References

- Aboulghar M, Baird DT, Collins J, Evers JLH, Fauser BCJM, Lambalk CB, Somigliana E, Sunde A, Crosignani PG, Devroey P *et al.* Intrauterine insemination. *Hum Reprod Update* 2009;**15**:265–277.
- Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr* 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, 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.
- Brandes M, Hamilton CJ, de Bruin JP, Nelen WL, Kremer JA. The relative contribution of IVF to the total ongoing pregnancy rate in a subfertile cohort. *Hum Reprod* 2010;**25**:118–126.
- Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008;**168**:656–664.

- Dreyer K, van Eekelen R, Tjon-Kon-Fat RI, van der Steeg JW, Steures P, Eijkemans M, van der Veen F, Hompes P, Mol B, van Geloven N. The therapeutic effect of hysterosalpingography in couples with unexplained subfertility: a post-hoc analysis of a prospective multi-centre cohort study. *Reprod Biomed Online* 2019;**38**: 233–239.
- Dreyer K, van Rijswijk J, Mijatovic V, Goddijn M, Verhoeve HR, van Rooij IAJ, Hoek A, Bourdrez P, Nap AWW, Rijnsaardt-Lukassen HGM et al. Oil-based or water-based contrast for hysterosalpingography in infertile women. *N Engl J Med* 2017;**376**:2043–2052.
- Farquhar CM, Liu E, Armstrong S, Arroll N, Lensen S, Brown J. Intrauterine insemination with ovarian stimulation versus expectant management for unexplained infertility (TUI): a pragmatic, open-label, randomised, controlled, two-centre trial. *Lancet* 2018;**391**: 441–450.
- Gran JM, Roysland K, Wolbers M, Didelez V, Sterne JA, Ledergerber B, Furrer H, von Wyl V, Aalen OO. A sequential Cox approach for estimating the causal effect of treatment in the presence of time-dependent confounding applied to data from the Swiss HIV Cohort Study. *Stat Med* 2010;**29**:2757–2768.
- Hunault CC, Habbema JD, Eijkemans MJ, Collins JA, Evers JL, Te Velde ER. Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod* 2004;**19**:2019–2026.
- Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005;**294**:218–228.
- Lawlor DA, Tilling KD, Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016;**45**:1866–1886.
- Moreau J, Gatimel N, Simon C, Cohade C, Lesourd F, Parinaud J, Léandri R. Potential chances for natural fertility influence results of intrauterine inseminations. *Eur J Obstet Gynecol Reprod Biol X* 2019;**4**:100058.
- Munafò MR, Smith GD. Repeating experiments is not enough. *Nature* 2018;**553**:399–401.
- Prentice RA, Gloeckler LA. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* 1978;**34**:57–67.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2017. <http://www.R-project.org/>.
- Steures P, van der Steeg JW, Hompes PG, Habbema JD, Eijkemans MJ, Broekmans FJ, Verhoeve HR, Bossuyt PM, van der Veen F, Mol BW. Intrauterine insemination with controlled ovarian hyperstimulation versus expectant management for couples with unexplained subfertility and an intermediate prognosis: a randomised clinical trial. *Lancet* 2006;**368**:216–221.
- Tjon-Kon-Fat RI, Bendsdorp AJ, Bossuyt PM, Koks C, Oosterhuis GJ, Hoek A, Hompes P, Broekmans FJ, Verhoeve HR, de Bruin JP et al. Is IVF-served two different ways-more cost-effective than IUI with controlled ovarian hyperstimulation? *Hum Reprod* 2015;**30**: 2331–2339.
- van der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Hompes PG, Broekmans FJ, van Dessel HJ, Bossuyt PM, van der Veen F, Mol BW. 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. Causal modelling in epidemiological practice. Ph.D. Thesis, 2011. *Chapter 8: Using Iterative Probability Weighting to Improve Causal Effect Estimates*. Amsterdam, the Netherlands: University of Amsterdam, 2011. <http://bit.ly/2kWMrRt>.
- 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* 2017a;**32**:346–353.
- van Eekelen R, van Geloven N, van Wely M, McLernon DJ, Eijkemans MJ, Repping S, Steyerberg EW, Mol BW, Bhattacharya S, van der Veen F. Constructing the crystal ball: how to get reliable prognostic information for the management of subfertile couples. *Hum Reprod* 2017b;**32**:2153–2158.
- van Eekelen R, van Geloven N, van Wely M, McLernon DJ, Mol F, Custers IM, Steures P, Bhattacharya S, Mol BW, van der Veen F et al. Is IUI with ovarian stimulation effective in couples with unexplained subfertility? *Hum Reprod* 2019;**34**:84–91.
- van Rijswijk J, van Welie N, Dreyer K, Pham CT, Verhoeve HR, Hoek A, De Bruin JP, Nap AWW, van Hooff MHA, Goddijn M et al. Tubal flushing with oil- or water-based contrast at hysterosalpingography for infertility: long-term reproductive outcomes of a randomized trial. *Fertil Steril* 2020;**114**:155–162.
- VanderWeele TJ. On the distinction between interaction and effect modification. *Epidemiology* 2009;**20**:863–871.
- Wang R, Danhof NA, Tjon-Kon-Fat RI, Eijkemans MJ, Bossuyt PM, Mochtar MH, van der Veen F, Bhattacharya S, Mol BWJ, van Wely M. Interventions for unexplained infertility: a systematic review and network meta-analysis. *Cochrane Database Syst Rev* 2019;**9**:CD012692.
- Wei L, Lin D, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc* 1989;**84**:1065–1073.