Cochrane Library

**Cochrane** Database of Systematic Reviews

# Individualised gonadotropin dose selection using markers of ovarian reserve for women undergoing in vitro fertilisation plus intracytoplasmic sperm injection (IVF/ICSI) (Review)

Lensen SF, Wilkinson J, Leijdekkers JA, La Marca A, Mol BWJ, Marjoribanks J, Torrance H, Broekmans FJ

Lensen SF, Wilkinson J, Leijdekkers JA, La Marca A, Mol BWJ, Marjoribanks J, Torrance H, Broekmans FJ.
Individualised gonadotropin dose selection using markers of ovarian reserve for women undergoing in vitro fertilisation plus intracytoplasmic sperm injection (IVF/ICSI).
*Cochrane Database of Systematic Reviews* 2018, Issue 2. Art. No.: CD012693.
DOI: 10.1002/14651858.CD012693.pub2.

**www.cochranelibrary.com**




WILEY

## TABLE OF CONTENTS

[Intervention Review]

# Individualised gonadotropin dose selection using markers of ovarian reserve for women undergoing in vitro fertilisation plus intracytoplasmic sperm injection (IVF/ICSI)

Sarah F Lensen[1], Jack Wilkinson[2], Jori A Leijdekkers[3], Antonio La Marca[4], Ben Willem J Mol[5], Jane Marjoribanks[1], Helen Torrance[3], Frank J Broekmans[3]

[1]Department of Obstetrics and Gynaecology, University of Auckland, Auckland, New Zealand. [2]Centre for Biostatistics, School of Health Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre (MAHSC), University of Manchester, Manchester, UK. [3]Department of Reproductive Medicine and Gynecology, University Medical Center, Utrecht, Netherlands. [4]Mother-Infant Department, University of Modena and Reggio Emilia, Clinica Eugin, Modena, Italy. [5]Discipline of Obstetrics and Gynaecology, School of Medicine, Robinson Research Institute, The University of Adelaide, Adelaide, Australia

Contact address: Sarah F Lensen, Department of Obstetrics and Gynaecology, University of Auckland, Park Rd, Grafton, Auckland, 1142, New Zealand. s.lensen@auckland.ac.nz.

**Editorial group:** Cochrane Gynaecology and Fertility Group.
**Publication status and date:** New, published in Issue 2, 2018.

**Citation:** Lensen SF, Wilkinson J, Leijdekkers JA, La Marca A, Mol BWJ, Marjoribanks J, Torrance H, Broekmans FJ. Individualised gonadotropin dose selection using markers of ovarian reserve for women undergoing in vitro fertilisation plus intracytoplasmic sperm injection (IVF/ICSI). *Cochrane Database of Systematic Reviews* 2018, Issue 2. Art. No.: CD012693. DOI: 10.1002/14651858.CD012693.pub2.



## ABSTRACT

### Background

During a cycle of in vitro fertilisation plus intracytoplasmic sperm injection (IVF/ICSI), women receive daily doses of gonadotropin follicle-stimulating hormone (FSH) to induce multifollicular development in the ovaries. Generally, the dose of FSH is associated with the number of eggs retrieved. A normal response to stimulation is often considered desirable, for example the retrieval of 5 to 15 oocytes. Both poor and hyper-response are associated with increased chance of cycle cancellation. Hyper-response is also associated with increased risk of ovarian hyperstimulation syndrome (OHSS). Clinicians often individualise the FSH dose using patient characteristics predictive of ovarian response such as age. More recently, clinicians have begun using ovarian reserve tests (ORTs) to predict ovarian response based on the measurement of various biomarkers, including basal FSH (bFSH), antral follicle count (AFC), and anti-Müllerian hormone (AMH). It is unclear whether individualising FSH dose based on these markers improves clinical outcomes.

### Objectives

To assess the effects of individualised gonadotropin dose selection using markers of ovarian reserve in women undergoing IVF/ICSI.

### Search methods

We searched the Cochrane Gynaecology and Fertility Group Specialised Register, Cochrane Central Register of Studies Online, MEDLINE, Embase, CINAHL, LILACS, DARE, ISI Web of Knowledge, ClinicalTrials.gov, and the World Health Organisation International Trials Registry Platform search portal from inception to 27th July 2017. We checked the reference lists of relevant reviews and included studies.

**Selection criteria**

We included trials that compared different doses of FSH in women with a defined ORT profile (i.e. predicted low, normal or high responders based on AMH, AFC, and/or bFSH) and trials that compared an individualised dosing strategy (based on at least one ORT measure) versus uniform dosing or a different individualised dosing algorithm.

**Data collection and analysis**

We used standard methodological procedures recommended by Cochrane. Primary outcomes were live birth/ongoing pregnancy and severe OHSS. Secondary outcomes included clinical pregnancy, moderate or severe OHSS, multiple pregnancy, oocyte yield, cycle cancellations, and total dose and duration of FSH administration.

**Main results**

We included 20 trials (N = 6088); however, we treated those trials with multiple comparisons as separate trials for the purpose of this review. Meta-analysis was limited due to clinical heterogeneity. Evidence quality ranged from very low to moderate. The main limitations were imprecision and risk of bias associated with lack of blinding.

**Direct dose comparisons in women according to predicted response**

All evidence was low or very low quality.

Due to differences in dose comparisons, caution is warranted in interpreting the findings of five small trials assessing predicted low responders. The effect estimates were very imprecise, and increased FSH dosing may or may not have an impact on rates of live birth/ongoing pregnancy, OHSS, and clinical pregnancy.

Similarly, in predicted normal responders (nine studies, three comparisons), higher doses may or may not impact the probability of live birth/ongoing pregnancy (e.g. 200 versus 100 international units: OR 0.88, 95% CI 0.57 to 1.36; N = 522; 2 studies; $I^2$ = 0%) or clinical pregnancy. Results were imprecise, and a small benefit or harm remains possible. There were too few events for the outcome of OHSS to enable any inferences.

In predicted high responders, lower doses may or may not have an impact on rates of live birth/ongoing pregnancy (OR 0.98, 95% CI 0.66 to 1.46; N = 521; 1 study), OHSS, and clinical pregnancy. However, lower doses probably reduce the likelihood of moderate or severe OHSS (Peto OR 2.31, 95% CI 0.80 to 6.67; N = 521; 1 study).

**ORT-algorithm studies**

Four trials compared an ORT-based algorithm to a non-ORT control group. Rates of live birth/ongoing pregnancy and clinical pregnancy did not appear to differ by more than a few percentage points (respectively: OR 1.04, 95% CI 0.88 to 1.23; N = 2823, 4 studies; $I^2$ = 34%; OR 0.96, 95% CI 0.82 to 1.13, 4 studies, $I^2$=0%, moderate-quality evidence). However, ORT algorithms probably reduce the likelihood of moderate or severe OHSS (Peto OR 0.58, 95% CI 0.34 to 1.00; N = 2823; 4 studies; $I^2$ = 0%, low quality evidence). There was insufficient evidence to determine whether the groups differed in rates of severe OHSS (Peto OR 0.54, 95% CI 0.14 to 1.99; N = 1494; 3 studies; $I^2$ = 0%, low quality evidence). Our findings suggest that if the chance of live birth with a standard dose is 26%, the chance with ORT-based dosing would be between 24% and 30%. If the chance of moderate or severe OHSS with a standard dose is 2.5%, the chance with ORT-based dosing would be between 0.8% and 2.5%. These results should be treated cautiously due to heterogeneity in the study designs.

**Authors' conclusions**

We did not find that tailoring the FSH dose in any particular ORT population (low, normal, high ORT), influenced rates of live birth/ongoing pregnancy but we could not rule out differences, due to sample size limitations. In predicted high responders, lower doses of FSH seemed to reduce the overall incidence of moderate and severe OHSS. Moderate-quality evidence suggests that ORT-based individualisation produces similar live birth/ongoing pregnancy rates to a policy of giving all women 150 IU. However, in all cases the confidence intervals are consistent with an increase or decrease in the rate of around five percentage points with ORT-based dosing (e.g. from 25% to 20% or 30%). Although small, a difference of this magnitude could be important to many women. Further, ORT algorithms reduced the incidence of OHSS compared to standard dosing of 150 IU, probably by facilitating dose reductions in women with a predicted high response. However, the size of the effect is unclear. The included studies were heterogeneous in design, which limited the interpretation of pooled estimates, and many of the included studies had a serious risk of bias.

Current evidence does not provide a clear justification for adjusting the standard dose of 150 IU in the case of poor or normal responders, especially as increased dose is generally associated with greater total FSH dose and therefore greater cost. However, a decreased dose in predicted high responders may reduce OHSS.

## PLAIN LANGUAGE SUMMARY

**Individualised stimulation dose using ovarian reserve markers in women doing in vitro fertilisation plus intracytoplasmic sperm injection (IVF/ICSI)**

**Background**

In planning an IVF cycle, doctors often decide the dose of stimulation drugs based on certain characteristics of each woman, such as their age. New tests have been developed that some specialists believe can better predict a woman's response to IVF stimulation. These are called ovarian reserve tests and are a general measure of the number of eggs available in the ovaries. It is unclear whether tailoring the doses of stimulation drugs based on the individual ovarian reserve tests can help to increase the chance of the woman getting pregnant and having a baby. It is also unclear whether the tests help to improve the safety of the IVF cycle, such as reducing the chances of a serious condition known as ovarian hyperstimulation syndrome (OHSS).

**Study characteristics**

We included two types of studies in this review. Direct dose comparison studies recruited women predicted to respond to IVF stimulation either poorly, normally, or excessively based on their ovarian reserve test. Researchers then randomly assigned these women to different doses of FSH to see whether the different doses would impact on IVF outcomes.

The ORT-algorithm studies divided a broader group of women into those whose stimulation dose was based on the women's ovarian reserve test and those receiving a standard dose of stimulation medication or a dose based on another characteristic about the women (other than their ovarian reserve).

In total we included 20 randomised controlled trials involving 6088 women.

**Key results**

***1. Direct dose comparison studies (low or very low quality evidence)***

In women predicted to respond poorly or normally to stimulation based on their ovarian reserve test, increasing the dose of stimulation medication did not seem to influence the chance of getting pregnant or having a baby, or the chance of OHSS. However, the included studies were small and compared different doses of medication. This made it difficult to say for sure that there is no difference between doses. For women predicted to respond poorly, if the chance of live birth with 150 IU is 11%, then the chance with 300/340 IU would be between 3.8% to 16%. For women predicted to have a normal response, if the chance of live birth or ongoing pregnancy with 150 IU is 19%, then the chance with 200/225 IU would be between 12% to 31%.

In women predicted to have an excessive response to stimulation, reducing the stimulation dose may or may not affect the chance of having a baby. If the chance of live birth with 100 IU is 26%, then the chance with 150 IU would be between 18% to 33%. However, it may reduce the rate of OHSS. If the chance of moderate or severe OHSS with a lower dose is 1.6%, then the chance with a higher dose would be between 1.3% and 9.6%.

***2. ORT-algorithm studies***

Moderate quality evidence from these studies suggested that using an ovarian reserve test to decide on the stimulation dose generally did not have much effect on the chance of getting pregnant and having a baby, but there could have been a relatively small difference one way or another. It did generally appear to reduce the chance of having OHSS when compared to giving all women the same dose of stimulation medication, but this evidence was low quality. Our findings suggest that if the chance of live birth with a standard dose were 26%, the chance with dosing based on an ovarian reserve test would be between 24% and 30% and that if the chance of moderate or severe OHSS with a standard dose were 2.5%, the chance with dosing based on an ovarian reserve test would be between 0.8% and 2.5%.

**Quality of the evidence**

We assessed the quality of the evidence as ranging from very low to moderate, due to limitations in study design (as researchers and participants often knew which treatment was assigned) and statistical imprecision, as the studies included too few women to provide meaningful results for the most important outcomes, such as having a baby.

## SUMMARY OF FINDINGS FOR THE MAIN COMPARISON [Explanation]

**ORT-based algorithm compared to standard dose of FSH for women undergoing IVF/ICSI**

**Patient or population**: women undergoing IVF/ICSI
**Setting**: hospital or fertility clinic
**Intervention**: ORT-based algorithm
**Comparison**: standard dose FSH

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Evidence summary |
|---|---|---|---|---|---|---|
| | Risk with 150 IU FSH | Risk with ORT-based algorithm | | | | |
| Live birth or ongoing pregnancy | 258 per 1000 | 266 per 1000 (235 to 300) | OR 1.04 (0.88 to 1.23) | 2823 (4 RCTs) | ⊕⊕⊕◯ Moderate[a] | Although the effect estimate remains imprecise, the pooled evidence suggests it is unlikely that ORT-algorithms impacted on rates of live birth or ongoing pregnancy |
| OHSS | *Severe* 8 per 1000 | 4 per 1000 (1 to 16) | OR 0.54 (0.14 to 1.99) | 1494 (3 RCTs) | ⊕⊕◯◯ Low[a,b] | Although the effect estimate remains imprecise, the pooled evidence suggests that ORT-algorithms reduce the incidence of OHSS by an unspecified amount |
| | *Moderate or severe* 25 per 1000 | 14 per 1000 (8 to 25) | OR 0.58 (0.34 to 1.00) | 2823 (4 RCTs) | | |
| Clinical pregnancy | 321 per 1000 | 313 per 1000 (280 to 349) | OR 0.96 (0.82 to 1.13) | 2823 (4 RCTs) | ⊕⊕⊕◯ Moderate[a] | Although the effect estimate remains imprecise, the pooled evidence suggests it is unlikely that ORT al- |

| | | | | |
|---|---|---|---|---|
| | | | | gorithms impacted on rates of clinical pregnancy |

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).

**CI**: confidence interval; **FSH**: follicle-stimulating hormone; **ICSI**: intracytoplasmic sperm injection; **IVF**: in vitro fertilisation; **OHSS**: ovarian hyperstimulation syndrome; **OR**: odds ratio; **ORT**: ovarian reserve test; **RCT**: randomised controlled trial.

**GRADE Working Group grades of evidence**

**High certainty**: we are very confident that the true effect lies close to that of the estimate of the effect.

**Moderate certainty**: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

**Low certainty**: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.

**Very low certainty**: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect

[a] Downgraded one level for serious risk of bias, associated mainly with performance bias due to lack of blinding and/or selective reporting.

[b] Downgraded one level for serious imprecision associated with small number of events.

# BACKGROUND

## Description of the condition

As many as 15% of couples experience difficulties getting pregnant and are defined as being subfertile (Thoma 2013). Treatments are available to help these couples conceive, such as intrauterine insemination, ovulation induction, in vitro fertilisation (IVF), and intracytoplasmic sperm injection (ICSI). IVF, with or without ICSI (referred to as IVF/ICSI when used together), is the leading treatment for most causes of infertility; however, the success rate remains modest at approximately 15% to 20% per cycle started and 30% per cumulative cycle (including fresh and all frozen embryo transfers) (Dyer 2016; Gunby 2010; Toftager 2017).

During an IVF/ICSI cycle, daily doses of the gonadotropin follicle-stimulating hormone (FSH) are used to induce multifollicular development in the ovaries. Generally the number of eggs retrieved depends on the dose of FSH; however, individual women's responses vary (Andersen 2006; Sunkara 2011). A low or poor ovarian response has been classified as the retrieval of three or fewer oocytes (Ferraretti 2011), and it often results in cycle cancellation, poor outcomes, and consequent stress and disappointment to the couple. The prevalence of poor response increases with age: approximately 10% to 15% of women aged 35 to 40 experience a poor response (Ferraretti 2011). Conversely, a hyper-response (or high response) is often defined as the retrieval of 15 to 20 or more oocytes and is associated with an exponential increase in the risk of ovarian hyperstimulation syndrome (OHSS) (Steward 2014; Youssef 2016). The incidence of OHSS is difficult to determine as there is no strict consensus definition (ASRM 2016). Historically, mild and moderate forms of OHSS were reasonably common, occurring in approximately 0% to 30% and 3% to 6% of cycles, respectively (Delvigne 2002). Severe OHSS is much less common but has potential to cause thromboembolic phenomena, multiple organ failure and even death (Delvigne 2002). More recent estimates of the incidence of moderate OHSS range from 0.6% to 5% per IVF/ICSI cycle (ASRM 2016; Calhaz-Jorge 2016; Kawwass 2015). Estimates of the rate of hospitalisation due to severe OHSS range from less than 0.01% to 0.3% of cycles (Kupka 2014; Harris 2016). This rate increases with the number of oocytes retrieved, reaching 4% with the retrieval of over 20 oocytes (Harris 2016).

The aim of most IVF cycles is to produce an embryo that leads to the live birth of a baby. Most specialists consider that obtaining a number of high-quality oocytes is an important step in this process and that the number of retrieved oocytes depends on many patient and treatment factors, two of which are the dose of FSH administered and the size of the pool of recruitable follicles. Up to certain limits, an increase in the FSH dose may increase the number of growing follicles and the resulting oocyte yields (Broer 2013b). As a consequence, the use of a very low dose of FSH may increase the risk of poor response. Conversely, a very high dose of FSH may increase the risk of hyper-response (in women with sufficient ovarian reserve). Although the retrieval of 5 to 15 eggs is correlated with the best chance of pregnancy and live birth (Sunkara 2011), it does not (necessarily) follow that increasing the FSH dose in order to obtain more eggs, for example in women with a previous or predicted poor response, will increase the probability of pregnancy for an individual woman.

## Description of the intervention

Clinicians may use a test of a woman's ovarian reserve to select the starting FSH dose for ovarian stimulation (Fauser 2017). This is done to reduce the variation in ovarian response. In general, this means administering higher doses to women with a low ovarian reserve test (ORT) result and lower doses to women with a high ORT result. Multifactorial dose-selection algorithms have also been developed, combining one or more ORT results with other patient characteristics such as age (La Marca 2012).

The oldest ovarian reserve test (ORT) is basal FSH (bFSH), measured in serum in the early follicular phase of a menstrual cycle. This was later supplemented by the antral follicular count (AFC) and more recently with anti-Müllerian hormone (AMH). AFC is measured by ultrasound and is a count of the number of antral follicles measuring about 2 mm to 10 mm (according to standard criteria) that are available in both ovaries (Broekmans 2010). It indicates the number of gonadotrophin-sensitive follicles available for stimulation in an IVF cycle. AMH is a protein expressed and secreted by the granulosa cells of the ovary and reflects the size of antral and pre-antral follicles (Visser 2006). AMH can be measured in serum and is a more direct and independent measure of the growing pre-antral and antral follicular pool (Seifer 2002; Van Rooij 2002).

## How the intervention might work

ORT-based individualisation of the FSH dose requires two components. First, there must be an ORT that can predict a woman's response when given a particular dose of FSH. Second, there must be a dose-response relationship between FSH and ovarian response, enabling manipulation of the response through adjustment of the dose administered.

In relation to the first component (prediction of response), diagnostic test studies have reported that ORT can be used to predict ovarian response to stimulation, with AMH and AFC being superior to bFSH (Broekmans 2006; Broer 2013a; Broer 2013b; Broer 2014; La Marca 2014). One meta-analysis of individual patient data found that for predicting excessive response, AMH and AFC showed similarly high performance (areas under the receiver operator characteristic curves (AUC) of 0.81 and 0.79, respectively) (Broer 2013a). However, bFSH had lower predictive value (AUC of 0.66). Predictive performance was improved by combining AMH and AFC (AUC 0.85). A second meta-analysis

showed that AFC and AMH as single tests both had high predictive value for poor response (AUC 0.78 and 0.76, respectively) and that combining these two tests did not substantially improve prediction (AUC 0.80, P = 0.19) (Broer 2013b).

In relation to the second component (dose-response relationship), a recent study indicated that increasing the dose of FSH increases oocyte yield in women with AMH between 5 pmol/L and 50 pmol/L (Arce 2014). For example, women who receive higher FSH doses will produce more follicles than those receiving lower FSH doses. However, the capacity to manipulate a woman's ovarian response may largely depend on her ovarian reserve. In particular, if a woman has relatively few antral follicles (and consequently is predicted to have a low ovarian response), then it may not be possible to compensate for this fact by increasing the FSH dose (Klinkert 2005; Lekamge 2008). It is important to remember that the relationship between the stimulation response and probability of pregnancy is poorly understood, so the use of surrogate outcomes such as number of eggs retrieved does not necessarily reveal the effects on pregnancy and live birth (Vail 2003). In fact, the above-mentioned individual patient data analysis found that ORTs did not improve prediction of ongoing pregnancy following IVF more than age alone (Broer 2013b).

### Why it is important to do this review

IVF/ICSI is expensive and invasive, and it requires extensive clinical monitoring. Those desiring a pregnancy often have to make a substantial financial investment, including time away from work, and the process is associated with a high emotional burden. If tailoring the dose of FSH can increase the likelihood of an appropriate response, it has the potential to increase pregnancy and live birth while reducing cancelled cycles (for either poor or hyper-response) and OHSS. Individualised FSH dosing also has the potential to be more cost-effective. On the other hand, an individualised approach to FSH dosing may be associated with greater cost in terms of price of FSH medication (if increased dose is recommended), cost of ORT testing itself, and increased administrative burden and complexity in monitoring of IVF cycles. However, there is no up-to-date review of the relevant literature.

## OBJECTIVES

To assess the effects of individualised gonadotropin dose selection using markers of ovarian reserve in women undergoing IVF/ICSI.

## METHODS

### Criteria for considering studies for this review

#### Types of studies

Published and unpublished randomized controlled trials (RCTs) were eligible for inclusion. We excluded non-randomised and quasi-randomised studies (e.g. studies with evidence of inadequate sequence generation such as allocation by alternate days or patient numbers), as they are associated with a high risk of bias (Higgins 2011). We did not use data from ongoing studies but will incorporate their results in future updates of the review. Cross-over trials would have been eligible, but as cross-over is not a valid design in the context of fertility trials, we would have considered only data from the first phase in meta-analyses.

Several trial designs are appropriate to the broad goal of investigating aspects of individualised FSH (Tajik 2013). Broadly, the two types of design included in this review are:

1. **direct dose comparison studies**, randomising women within a given ORT range to one of several doses of FSH; and
2. **ORT-algorithm studies**, randomising women either to dose selection according to their ORT value using an algorithm, or to dose selection without ORT or using an alternative algorithm.

The first type of design (direct dose comparison studies) allocates women of a given ORT profile to one of two (or more) doses of FSH, in order to compare the responses of similar women under each of the doses. An example would be a trial of women with low AMH (predicted low responders) who are randomized to two different doses of FSH (e.g. 150 international units (IU) vs 300 IU). This type of design is useful for establishing whether there is a dose-response relationship between FSH and outcome in subgroups of women, or for identifying the optimal FSH dose for women with a given set of predictive characteristics. This design is able to tell us whether certain groups of women would benefit from a particular FSH dose (Tajik 2013). We will use the terms low and high responders to refer to the *predicted* response of women, and the terms poor and hyper-response to refer to the *observed* response of women to ovarian stimulation, usually measured by the number of oocytes retrieved.

The second type of design (ORT-algorithm studies) randomises women either to FSH dose selection determined by an algorithm including ORT, or to a standard FSH dose (i.e. for all women regardless of their ORT). In this design, all women in the control arm receive the same dose of FSH, and women in the intervention arm receive different doses of FSH according to their individual characteristics, such as AMH level. A variant of this type of design randomises women to one of two (or more) individualised dose-selection algorithms/policies, where the comparator algorithms may or may not include ORT. The purpose of designs of this type is to compare an ORT-individualised dose-selection algorithm versus either a uniform dose or alternative dose-selection policy.

We included studies of both design types in separate comparisons in this review. Sometimes, trials were not explicitly presented as falling into one of the above types of design, but nonetheless it was possible to interpret and analyze them in such a way that they were

equivalent. In these cases, the trials were eligible for this review.

## Types of participants

### Direct dose comparison studies

For these studies to be eligible, the study population had to be women undergoing IVF/ICSI, categorised as either predicted low, normal or high responders based on at least one ORT (AMH, bFSH, or AFC) (or providing data that enabled categorisation by review authors). Studies including unselected populations were not eligible unless we could obtain data from eligible subgroups within the studies.

### ORT-algorithm studies

Studies of this type had to include a (possibly unselected) population of women undergoing IVF/ICSI.

Studies in women who did not plan to undergo embryo transfer, for example women planning oocyte donation or fertility preservation, or who were receiving donated oocytes, were excluded. We excluded studies including only women with polycystic ovarian syndrome (PCOS), which represents a distinct clinical entity and likely warrant unique individualised dosing algorithms. We included studies including only some women with PCOS and attempted to obtain the data excluding them. There were no exclusion criteria related to age, cause of infertility, or previous IVF/ICSI exposure.

## Types of interventions

### Included interventions

Studies comparing ovarian stimulation doses with each other (direct dose comparison studies) or comparing ORT-based FSH dose individualisation versus an alternative dosing policy (ORT-algorithm studies) were eligible for inclusion. Eligible individualised policies include those where the dose was selected, at least in part, using the woman's ORT measure (e.g. AMH, AFC, bFSH). We also included policies of dose selection on the basis of combinations of characteristics, provided one or more ORTs were amongst the considered factors. Studies comparing doses of human menopausal gonadotropin (HMG), which contains both FSH and luteinising hormone, were also eligible.

Additionally, we included ORT-algorithm studies comparing different preparations and brands, provided that the dose-selection algorithm varied between study arms. This reflected the more pragmatic nature of the questions being answered by these designs. Studies that allowed dose adjustment following a certain number of days of administration of the randomized dose were eligible as long as that adjustment was permitted in both study arms. This was subject to sensitivity analysis.

### Excluded interventions

For direct dose comparison studies, we excluded studies comparing different preparations, brands, or routes of administration, since treatment effects in these studies might not be attributable to differences in dose.

We excluded:

- studies comparing HMG to pure FSH preparations;
- studies using medications other than gonadotropins, such as clomiphene citrate or letrozole;
- studies comparing doses of corifollitropin alfa;
- studies comparing step-up/step-down protocols, or protocols amending the FSH dose in only one arm after commencing stimulation, for example coasting or withholding FSH for a number of days; and
- studies comparing different stimulation regimens (for example, GnRH agonist versus GnRH antagonist).

## Types of outcome measures

### Primary outcomes

1. Live birth or ongoing pregnancy. Ongoing pregnancy was defined as evidence of a gestational sac with fetal heart motion at or after twelve weeks' gestation, confirmed with ultrasound (Harbin Consensus Workshop Group 2014). Ongoing pregnancy data were only used when live birth data were not available. In the event that studies included multiple cycles for an individual woman, we also reported cumulative live birth. If studies reported the live birth outcome of the fresh transfer and the first frozen transfer for women with freeze-all cycles, we also reported this outcome separately. We counted multiple live births (e.g. twins or triplets) as one live birth event.
2. Severe ovarian hyperstimulation syndrome (OHSS) (as defined by authors).

### Secondary outcomes

1. Clinical pregnancy, defined as evidence of an intrauterine gestational sac on ultrasound or other definitive signs of pregnancy, including ectopic pregnancy.
2. Time to clinical pregnancy.
3. Moderate or severe OHSS (as defined by study authors).
4. Multiple pregnancy in randomized women.
5. Multiple pregnancy in women with clinical pregnancy, noting that this does not reflect a randomized comparison.
6. Number of oocytes retrieved per woman randomized.
7. Poor response to stimulation (as defined and prespecified by trial authors).
8. Normal response to stimulation (as defined and prespecified by trial authors).
9. Hyper-response to stimulation (as defined and prespecified by trial authors).

10. Cycle cancellations for hyper-response (including freeze-all cycles).
11. Cycle cancellations for poor response.
12. Cycle cancellations for poor or hyper-response.
13. Women with at least one transferable embryo.
14. Total dose of FSH.
15. Duration of FSH administration.
16. Cost per woman randomized.
17. Cumulative live birth rate.

## Search methods for identification of studies

We searched for all published and unpublished RCTs that met our inclusion criteria, without language or date restriction and in consultation with the Cochrane Gynaecology and Fertility Group (CGF) Information Specialist.

### Electronic searches

We searched the following electronic databases, trial registers and websites in November 2016 and on the 27th July 2017.

- The Gynaecology and Fertility Group (CGF) Specialised Register of Controlled Trials (searched 27th July 2017) (Appendix 1).
- The Cochrane Central Register of Studies Online (CRSO) (searched 27th July 2017) (Appendix 2).
- MEDLINE (from 1946 to 27th July 2017) (Appendix 3).
- Embase (from 1980 to 27th July 2017) (Appendix 4).
- CINAHL (from 1961 to 27th July 2017) (Appendix 5).

We combined the MEDLINE search with the Cochrane highly sensitive search strategy for identifying randomized trials, described in section 6.4.11 of the *Cochrane Handbook of Systematic Reviews of Interventions* (Higgins 2011). We combined the Embase and CINAHL searches with trial filters developed by the Scottish Intercollegiate Guidelines Network (SIGN) (www.sign.ac.uk/methodology/filters.html#random).
We searched other electronic sources of trials from their inception to 27th July 2017.

- Trial registers for ongoing and registered trials: www.clinicaltrials.gov (a service of the US National Institutes of Health) and www.who.int/trialsearch/Default.aspx (the World Health Organisation International Trials Registry Platform search portal).
- DARE (Database of Abstracts of Reviews of Effects) on the Cochrane Library: onlinelibrary.wiley.com/o/cochrane/cochrane_cldare_articles_fs.html (for reference lists from relevant non-Cochrane reviews);
- The Web of Knowledge: wokinfo.com/ (another source of trials and conference abstracts).
- OpenGrey: www.opengrey.eu/ for unpublished literature from Europe.
- LILACS database: regional.bvsalud.org/php/index.php?lang=en.
- PubMed and Google Scholar (for recent trials not yet indexed in the major databases).

We detail the search strategies used in the Appendices.

### Searching other resources

We handsearched reference lists of articles retrieved by the search and contacted experts in the field to obtain additional data. We also handsearched relevant journals and conference abstracts that were not covered in the CGF register, in liaison with the Information Specialist.

## Data collection and analysis

### Selection of studies

After an initial screen of titles and abstracts retrieved by the search, we retrieved the full texts of all potentially eligible studies. Two review authors independently examined these full-text articles for compliance with the inclusion criteria and selected studies eligible for inclusion in the review. We corresponded with study investigators as required to clarify study eligibility. We resolved disagreements as to study eligibility by discussion or by involving a third review author. We documented the selection process with a PRISMA flow chart.

### Data extraction and management

Two review authors independently extracted data from eligible studies using a data extraction form designed and pilot-tested by the authors. We resolved any disagreements by discussion or by involving a third review author. Data extracted included study characteristics and outcome data. Where studies had multiple publications, we collated the multiple reports; the study rather than the report was the unit of interest in the review. Studies with multiple reports had a single study ID with multiple references.
We corresponded with study investigators for further information on methods, results or both, as required.

### Assessment of risk of bias in included studies

Two review authors independently assessed the included studies for risk of bias using the Cochrane 'Risk of bias' assessment tool, which considers bias arising from: selection (random sequence generation and allocation concealment), performance (blinding of participants and personnel), detection (blinding of outcome assessors), attrition (incomplete outcome data), reporting (selective reporting), and other causes (Higgins 2011). We resolved any disagreements by discussion or by involving a third review author.

We described all judgements fully and presented the conclusions in the 'Risk of bias' table, which we incorporated into the interpretation of the review findings both qualitatively and formally, by means of sensitivity analyses. Where identified studies failed to report the primary outcome of live birth but did report interim outcomes such as pregnancy, we undertook informal assessment as to whether the interim values (e.g. pregnancy rates) were similar to those reported in studies that also reported live birth.

We considered the following methods of random sequence generation adequate.

- Referring to a random number table.
- Using a computer random number generator.
- Coin tossing.
- Shuffling cards or envelopes.
- Throwing dice.
- Drawing of lots.

We considered it insufficient to state that the study was 'randomized' and rated these studies at unclear risk of bias.

We considered the following methods of allocation concealment adequate.

- Central allocation (including telephone, Internet-based and pharmacy-controlled randomization).
- Sequentially numbered, opaque, sealed envelopes (SNOSE).

We considered blinding of participants and personnel to carry a low risk of bias if there was a description of adequate blinding measures, for example administering doses that were identical in appearance. There was potential for performance bias, as some methods and outcomes were not strictly objective, such as cycle cancellation for poor or hyper-response, number of eggs collected, embryo selection for embryo transfer, decision to freeze all embryos, etc. Additionally, in trials that allowed dose adjustment during stimulation, there was potential for performance bias, so we considered the risk to be high in these cases.

We considered the domain 'Blinding of outcome assessors' to be relevant only for OHSS outcomes, and we rated it as low risk for other outcome variables. This is because diagnosis and classification of OHSS can be subjective. For OHSS outcomes, we rated the domain as being at low risk of bias if there was some description of adequate blinding measures. For example, if the text stated that diagnosis of OHSS was done by a clinician not involved in the trial, we rated the risk of bias as low for this domain.

We considered studies with a loss to follow-up of 15% or more as being at high risk of attrition bias. This cutoff is arbitrary, but there is value in prespecifying a criterion in order to reduce post hoc decisions.

We considered studies that had collected more outcome measures than were reported in the paper as being at high risk of reporting bias. It was often difficult to determine which outcomes they measured unless a study protocol was available. Therefore, in the absence of a protocol, we might have rated the risk of bias as unclear. However, if a study reported all expected outcomes, we assigned a low risk rating.

### Measures of treatment effect

For dichotomous data (e.g. live birth rates), we used the numbers of events in the control and intervention groups of each study to calculate Mantel-Haenszel odds ratios (ORs). If event rates in a particular analysis were low, however, we preferred Peto's method (e.g. multiple pregnancy and OHSS). For continuous data (e.g. total dose of FSH), if all studies reported exactly the same outcomes we calculated the mean difference (MD) between treatment groups. Had studies reported time-to-event data, we would have used hazard ratios (HRs) as the measure of treatment effect.

We reversed the direction of effect of individual studies to ensure consistency across trials (for example, in direct dose comparison studies, consistently ordering the higher and lower doses). We presented 95% confidence intervals (CIs) for all outcomes. Where data to calculate ORs or MDs were not available, we utilized the data available to proceed with the most reasonable analysis available (e.g. test statistics, P values). We emphasised the magnitude, precision, and direction of effects rather than relying on arbitrary and uninformative standards of statistical significance.

### Unit of analysis issues

We performed the primary analysis with the denominator of randomized women; we also included per pregnancy data for the outcome of multiple pregnancy, as this better reflects the proportion of pregnancies that were multiple, but readers should interpret these results with caution, as they do not represent a randomized comparison. For time to clinical pregnancy, we had anticipated that the unit of time in the analysis would have been the cycle; however, no study reported this outcome (two trials reported time to ongoing pregnancy only: Oudshoorn 2017; Van Tilborg 2017). We summarized in narrative data that did not allow valid analyses. Where studies followed up women over multiple treatment cycles, we included 'cumulative' birth events in the numerator as a separate outcome.

### Dealing with missing data

For all outcomes, we carried out analyses on an intention-to-treat basis as far as possible, that is, we attempted to analyze all participants in the group to which they were randomized, regardless of whether or not they received the allocated intervention. The denominator for each outcome was the number randomized, except for the outcome 'multiple pregnancy in women with clinical pregnancy'. In relation to the primary outcome live birth, we assumed that those who dropped out of the study did not have a successful treatment outcome. When necessary, we contacted the authors of included studies to obtain missing data.

Additional statistical analyses were required for an intention-to-treat analysis of the outcome 'number of oocytes'. It was common for studies to exclude cycles cancelled when reporting this outcome. This is akin to active censoring, which violates the randomization in the study and biases the estimated treatment effect. In these cases, we recalculated the mean numbers of oocytes including all randomized participants by setting the values for participants with cancelled cycles to zero and including these women in the divisor. We also had to impute the corresponding standard deviations, which would be larger than those calculated using only uncensored patients. On the basis of simulations, we determined that adding half the difference between the reported and the new mean to the reported standard deviation produced a suitably adjusted estimate. This amounted to a small adjustment (less than one oocyte). These imputed standard deviations have the disadvantage of probably being wrong, but the advantage of being an improvement over the reported values. For one study (Tasker 2010), individual patient data were available, allowing us to conduct multiple imputation for any cancelled cycles. Some studies reported the median rather than the mean number of oocytes. Because the distribution of numbers of oocytes is skewed, we imputed the mean by adding one to the median. This small adjustment was deemed to be appropriate on the basis of analyses conducted using the Tasker 2010 data. Finally, the skewed distribution meant that meta-analysis based on an assumption of a normal distribution was not appropriate. Accordingly, we adopted a method for the meta-analysis of skewed data (method 1 in Higgins 2008). Briefly, this involves approximating the difference in log scale means and a corresponding standard error, based on the summary data available. These were synthesised using the generic inverse variance functionality in RevMan. For these reasons, the mean differences reported here differ slightly from those in the papers.

### Assessment of heterogeneity

We considered whether the clinical and methodological characteristics of the included studies were sufficiently similar for meta-analysis to provide a clinically meaningful summary. We assessed statistical heterogeneity by the measure of the $I^2$. We interpreted an $I^2$ measurement greater than 50% as indicating substantial heterogeneity (Higgins 2003), although we acknowledge that this threshold is essentially arbitrary.

### Assessment of reporting biases

In view of the difficulty of detecting and correcting for publication bias and other reporting biases, we aimed to minimise their potential impact by ensuring a comprehensive search for eligible studies and by being alert for duplication of data. If there had been 10 or more studies in an analysis, we would have used a funnel plot to explore the possibility of small study effects (a tendency for estimates of the intervention effect to be more beneficial in smaller studies).

### Data synthesis

Although we had anticipated that the included studies would display considerable protocol heterogeneity, the data synthesis scheme we had proposed in the review protocol could not comfortably accommodate the variety of eligible direct dose comparison studies we identified in the search. Accordingly, we modified it, and the modified scheme we eventually used is described here (see also Differences between protocol and review).

For direct dose comparison studies (women with a given ORT measurement randomized to one of several doses), we considered the following comparisons.

- **Comparison 1**. All pairwise dose comparisons tested in women predicted to have a low response on the basis of one or more ORT.
- **Comparison 2**. All pairwise dose comparisons tested in women predicted to have a normal response on the basis of one or more ORT.
- **Comparison 3**. All pairwise dose comparisons tested in women predicted to have a high response on the basis of one or more ORT.

We made a post hoc decision to pool studies within each predicted response category (low, normal, high) if they shared the same comparator dose (e.g. to pool a trial comparing 200 IU vs 150 IU with another trial comparing 300 IU vs 150 IU). We made this decision in the final stages of the review after observing that most of the included studies compared different dose sets. This pooling, to the extent that it is interpretable, answered a broader question of the data: compared to a dose of 150 IU, does a higher dose offer any benefit in women with predicted low-response? Readers should consider these pooled comparisons as summaries of the studies, rather than as unified estimates of underlying treatment effects.

We used the following cutoffs to guide the categorisation of women as required based on categorisations used previously (e.g. Arce 2014; Jayaprakasan 2010; Oudshoorn 2017).

- AMH < 7 pmol/L, AFC < 7, bFSH > 10 IU/L categorised as predicted low responders.
- AMH 7 pmol/L to 21 pmol/L, AFC 7 to 15 categorised as predicted normal responders (bFSH is not considered to be reliable predictor for normal response).
- AMH > 21 pmol/L, AFC > 15 categorised as predicted high responders (bFSH is not considered to be reliable predictor for hyper-response).

We considered the ORT values of the cohorts in each study as a potential source of heterogeneity but determined that it would not be feasible to stratify the trials further on the basis of type of ORT.

In the review protocol, we noted that it was not possible to anticipate the combinations of study arms that would be compared in ORT-algorithm studies. Accordingly, we modified the basic scheme we had proposed in the protocol to accommodate the eligi-

ble trials we found in the search (see Differences between protocol and review) and presented the modified scheme we used here.
For ORT-algorithm studies (women randomized to either have a dose selected according to their ORT value using an algorithm, or to a uniform dose/dose selected using an alternative algorithm), we considered the following comparisons.

- **Comparison 4**. ORT-based dose selection algorithm for ovarian stimulation vs dose selection without ORT (including uniform dosing policies).
- **Comparison 5**. ORT-based dose selection algorithm for ovarian stimulation vs alternative ORT-based dose selection algorithm.

Within comparison 4, we stratified the trials according to the comparator arm and did not consider it to be meaningful to pool across strata. Specifically, we deemed it appropriate to pool studies comparing ORT-based algorithms to a uniform dose if that dose was the same in the different studies. We did not pool studies with non-ORT dose selection algorithms as comparator interventions with the studies with uniform dose control groups, however. We would stress that pooled estimates derived from comparison 4 should be considered as summaries of the effects estimated in the included studies, rather than an estimate of a distinct underlying treatment effect.
The trials included in comparison 5 each made a unique comparison between ORT-based algorithms, and we did not consider it appropriate to pool these studies.
Any increase in the odds of a particular outcome under a higher dose (for direct dose comparison studies) or under an ORT-based algorithm (ORT-algorithm studies), regardless of whether it was beneficial (e.g. live birth) or detrimental (e.g. adverse effects), was displayed graphically in the meta-analyses to the right of the centre line. Any decrease in the odds of an outcome was displayed to the left. For comparison 5, comparing ORT-based algorithms against one another, the decision of which algorithm to treat as the comparator and which to treat as the 'experimental' treatment was essentially arbitrary.
When trials reported outcomes for total dose of FSH and duration of FSH as medians, we treated these as means, assuming a symmetrical distribution; however, this assumption will be poor.

### Subgroup analysis and investigation of heterogeneity

We intended to conduct subgroup analyses where at least one trial fitted within each subgroup, data were available and substantial heterogeneity existed, to determine the separate evidence within the following subgroups for primary outcomes only.

1. Predicted response category (e.g. high responders, normal responders, low responders). The stratification of women into predicted response categories was already a feature of our analysis plan for direct dose comparison studies. However, we intended to consider the evidence, where available, for subgroups determined by predicted response category in ORT-algorithm studies.
2. Age (less than 35 years, 35 to 40 years, more than 40 years)
3. IVF protocol type (e.g. long GnRH agonist, short GnRH agonist (or 'Flare'), antagonist)

Where we detected substantial heterogeneity, we explored possible explanations in sensitivity analyses. We incorporated statistical heterogeneity into our interpretation of results, paying particular attention to any variation in the direction of effect.

### Sensitivity analysis

We planned to conduct sensitivity analyses for the primary outcomes to determine whether the conclusions were robust to arbitrary decisions made regarding the eligibility and analysis. These analyses included consideration of whether the review conclusions would have differed if:

1. eligibility had been restricted to studies at low risk of bias (defined as studies rated as being at low risk of bias with respect to sequence generation and allocation concealment, and not rated as at high risk of bias in any of the domains assessed);
2. a random-effects model had been adopted;
3. ongoing pregnancy data were not combined with live birth data; or
4. studies that allowed dose adjustment.

### Summary of findings table

We prepared 'Summary of findings' tables using GRADEpro software and Cochrane methods (GRADEpro GDT 2014; Higgins 2011). These tables evaluated the overall quality of the body of evidence for the main review outcomes (live birth or ongoing pregnancy, OHSS, clinical pregnancy) in each of the main comparisons of the review, using GRADE criteria. There was one comparison for each patient subgroup in the direct dose comparison studies (predicted low responders, normal responders, high responders) and a further comparison for use of ORT-based algorithms versus dosing without ORT. GRADE criteria relate to study limitations (i.e. risk of bias), inconsistency of effect, imprecision, indirectness and publication bias. Two review authors independently made judgements on evidence quality (high, moderate, low, or very low), resolving disagreements by discussion. We justified, documented, and incorporated these judgements into the reporting of results for each outcome.
We extracted study data, formatted our comparisons in data tables and prepared a 'Summary of findings' table before writing the results and conclusions of the review.

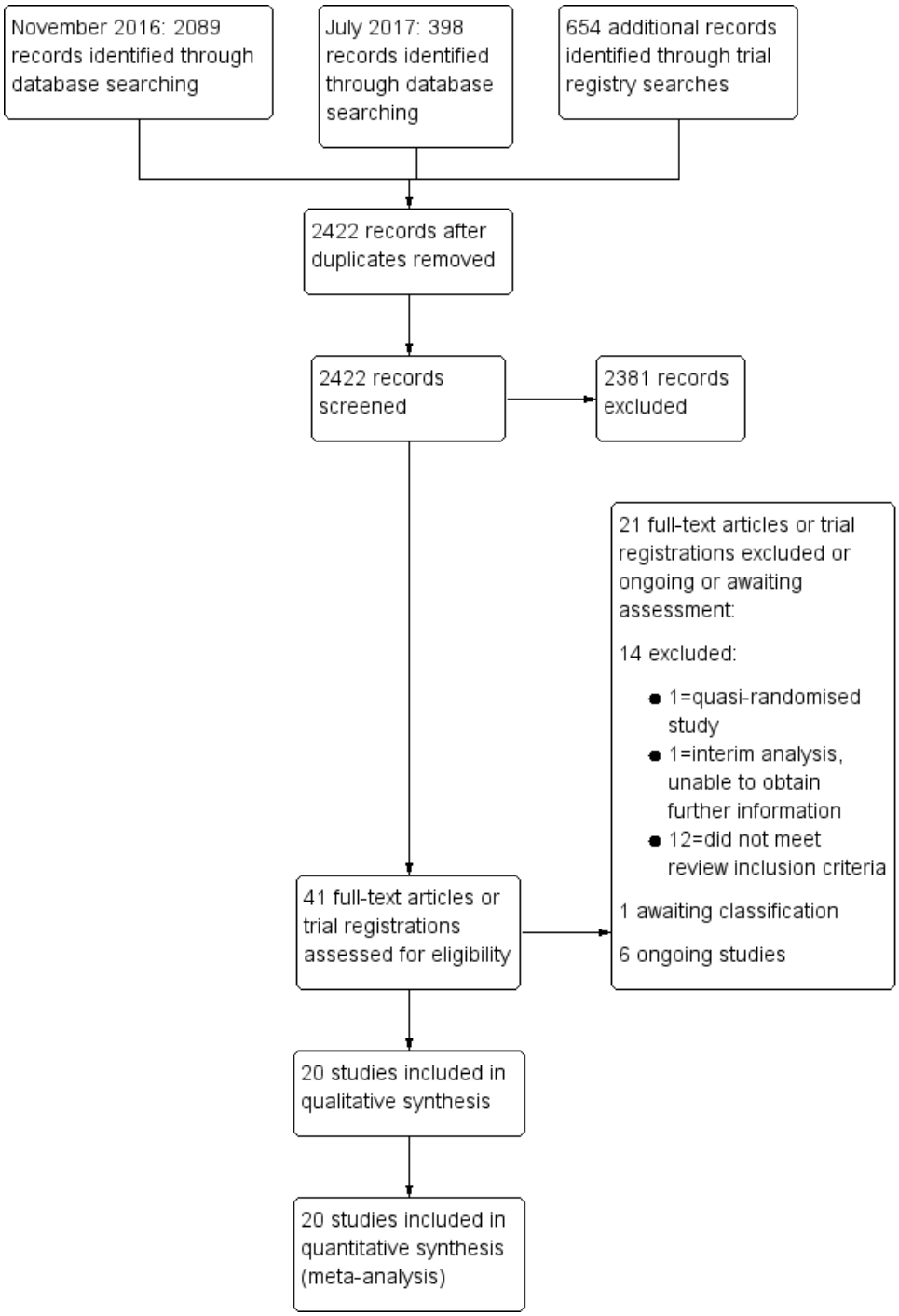
## RESULTS

### Description of studies

### Results of the search

Our searches yielded 2422 unique articles (Figure 1). We excluded 2381 records based on screening the title and abstract and retrieved the full text of 41 records for more detailed assessment. We excluded 21 articles, mostly because they did not meet review criteria. Among the trials excluded from the review were six ongoing studies, whose status trial investigators confirmed in four cases (EUCT2012-004969-40; NCT02430740; NCT02739269; Singh 2015). We were unable to contact the investigators of NCT01794208 or to ascertain the status of CTRI/2016/10/007367 from the investigators.

**Figure 1. Study flow diagram.**



We included 20 studies in the review, including 3 that we essentially treat as multiple trials for the purposes of this review (Arce 2014; Harrison 2001; Van Tilborg 2017). Most studies were published as full-text articles; however, one study was available as an abstract only, and we obtained the individual participant data from the trialists to enable further data analysis (Tasker 2010).

## Included studies

### Study design and setting

We included 20 parallel-design randomized controlled trials in the review. Seventeen studies had two arms (Allegra 2017; Bastu 2016; Cavagna 2006; Hoomans 2002; Jayaprakasan 2010; Klinkert 2005; Lan 2013; Lefebrve 2015; Magnusson 2017; Nyboe Andersen 2017; Olivennes 2015; Oudshoorn 2017; Out 2004; Popovic-Todorovic 2003; Tan 2005; Tasker 2010; YongPYK 2003), two studies had four arms (Harrison 2001; Van Tilborg 2017), and one had five arms (Arce 2014). Two studies had additional trial arms that were not relevant and which we excluded from this review (Arce 2014; Bastu 2016).

Most studies took place in European countries, including Denmark (Popovic-Todorovic 2003), Ireland (Harrison 2001), Italy (Allegra 2017), the Netherlands (Klinkert 2005; Oudshoorn 2017; Van Tilborg 2017), the UK (Jayaprakasan 2010; Out 2004; Tasker 2010; YongPYK 2003), and Sweden (Magnusson 2017). Three studies were conducted across multiple European countries (Arce 2014; Nyboe Andersen 2017; Olivennes 2015). Two studies took place in Canada (Lefebrve 2015; Tan 2005), and there was one study each from Brazil (Cavagna 2006), Turkey (Bastu 2016), and Vietnam (Lan 2013), along with one in multiple Asian countries (Hoomans 2002). Eleven studies took place in a single centre (Allegra 2017; Bastu 2016; Cavagna 2006; Harrison 2001; Jayaprakasan 2010; Klinkert 2005; Lan 2013; Lefebrve 2015; Magnusson 2017; Tasker 2010; YongPYK 2003), and nine were multicentre (Arce 2014; Hoomans 2002; Nyboe Andersen 2017; Olivennes 2015; Oudshoorn 2017; Out 2004; Popovic-Todorovic 2003; Tan 2005; Van Tilborg 2017).

Two of the direct dose comparison studies were conducted in tandem as part of a wider cohort study (Oudshoorn 2017; Van Tilborg 2017). One of these studies is essentially treated as two separate trials for the purpose of this review under the same reference in different comparisons (Van Tilborg 2017). Further, these trials are all merged to produce one ORT-algorithm study (Oudshoorn 2017).

### Participants and interventions

All studies but Harrison 2001 had inclusion criteria based on age. Most studies used a long agonist protocol; however, four used an antagonist protocol (Arce 2014; Bastu 2016; Nyboe Andersen 2017; Out 2004), one used a microdose flare protocol (Lefebrve 2015), and two did not require the use of any specific stimulation protocol (Oudshoorn 2017; Van Tilborg 2017). Ten studies permitted dose adjustment during the stimulation phase (Allegra 2017; Harrison 2001; Klinkert 2005; Lan 2013; Magnusson 2017; Nyboe Andersen 2017; Olivennes 2015; Out 2004; Popovic-Todorovic 2003; Tan 2005), while six studies did not permit adjustment for any reason (Arce 2014; Bastu 2016; Cavagna 2006; Hoomans 2002; Jayaprakasan 2010; Lefebrve 2015); in one study it was unclear (Tasker 2010). Two studies permitted dose-adjustment only between IVF cycles (Oudshoorn 2017; Van Tilborg 2017), which is only relevant for the outcome of cumulative live birth rate reported in this study (over 18 months).

#### Direct dose comparison studies

All 13 direct dose comparison studies (including three studies that are used twice in different comparisons/subgroups) focused on a population defined as either predicted low, normal, or high responders based on at least one ORT measure (AMH, AFC or bFSH), or reported on at least one of these measures demographically (as per the review protocol). Five studies involved predicted low responders (Bastu 2016; Harrison 2001; Klinkert 2005; Lefebrve 2015; Van Tilborg 2017); nine studies, predicted normal responders (Arce 2014; Cavagna 2006; Harrison 2001; Hoomans 2002; Jayaprakasan 2010; Out 2004; Tan 2005; Van Tilborg 2017; YongPYK 2003); and two studies, predicted high responders (Arce 2014; Oudshoorn 2017). Of the ORTs, six used AMH to define their population or reported AMH as a demographic (Arce 2014; Bastu 2016; Jayaprakasan 2010; Lefebrve 2015; Oudshoorn 2017; Van Tilborg 2017), seven used or reported AFC (Arce 2014; Bastu 2016; Klinkert 2005; Jayaprakasan 2010; Lefebrve 2015; Oudshoorn 2017; Van Tilborg 2017), and all but two used or reported bFSH (Oudshoorn 2017; Van Tilborg 2017). There was significant variation in the thresholds and application of ORT as eligibility criteria. For example, some trials required participants to satisfy all ORT criteria to be eligible (e.g. Jayaprakasan 2010 required participants to have bFSH of less than 12 IU/L and AFC 8 to 21), and other trials permitted participants to satisfy at least one of a number of criteria (e.g. Bastu 2016 required participants to meet at least two of the three following criteria: age over 40 years, previous poor response, abnormal ORT measure).

Each of the five studies in **low responders** employed a separate comparison, and we pooled these as follows.

- 300/400 IU vs 150 IU: 300 IU vs 150 IU (Klinkert 2005), 450 IU vs 150 IU (Van Tilborg 2017).
- 400/450 IU vs 300 IU: 400 IU vs 300 IU (Harrison 2001), 450 IU vs 300 IU (Bastu 2016).

- 600 IU vs 450 IU: (Lefebrve 2015).

There were three separate pooled comparisons among the nine studies in predicted **normal responders**.

- 200 vs 100 IU (Hoomans 2002; Tan 2005).
- 225/200 IU vs 150 IU (Cavagna 2006; Harrison 2001; Out 2004; Van Tilborg 2017; YongPYK 2003).
- 300 vs 225 IU (Jayaprakasan 2010).

A five-arm dosing study used a novel FSH (FE 999049), expressed in µg rather than IU, which is not directly translatable to IU. Therefore, we were unable to pool the data from this trial with the other studies and instead present the five dosing arms in separate forest plots for each outcome in incremental comparisons (i.e. 5.2 µg versus 6.9 µg, 6.9 µg versus 8.6 µg, 8.6 µg, versus 10.3 µg, 10.3 µg versus 12.1 µg).

The study with five arms also had a strata of women included in the comparison for predicted **high responders** (Arce 2014), along with a second study (Oudshoorn 2017). In total, the 13 direct dose comparison studies included 752 low responders, 1774 normal responders, and 618 high responders.

#### ORT-algorithm studies

There were eight ORT-algorithm studies included, which generally recruited women of a broader ORT spectrum. We merged the data from two of the direct dose comparison studies conducted in tandem as part of a wider cohort study, Oudshoorn 2017 and Van Tilborg 2017, and included them as one ORT-algorithm study (Oudshoorn 2017). All eight studies used or reported AFC, all but two also used or reported AMH (Popovic-Todorovic 2003; Oudshoorn 2017), and all but two also used bFSH (Magnusson 2017; Oudshoorn 2017).

Five studies compared an ORT-based algorithm to a method that did not use any ORT, either a standard dose of 150 IU (Nyboe Andersen 2017; Olivennes 2015; Oudshoorn 2017; Popovic-Todorovic 2003), or an algorithm not using ORT (Allegra 2017). In this latter study, the dose selection in the non-ORT arm was based solely on age (women aged 35 years or less received 150 IU, those aged more than 35 years old received 225 IU). Three studies compared two different ORT-based algorithms with each other. One study compared an AMH-based algorithm versus an AFC-based algorithm (Lan 2013), one study compared an AFC-based algorithm versus an algorithm using both AFC and AMH (Magnusson 2017), and one study compared an algorithm based on a number of markers (age, bFSH, oestradiol, and polycystic ovaries status) versus an algorithm based on AMH and AFC in addition to the other markers (Tasker 2010). In total, the eight ORT-algorithm studies included 3888 participants, 3017 of whom contributed to a comparison between an ORT-algorithm and a non-ORT method of dose selection, and 871 to a comparison of two different ORT-based algorithms.

### Excluded studies

We excluded 14 studies, 13 of which did not measure or report at least one of AMH, bFSH, AFC, and another that we discovered had been quasi-randomised following author correspondence (Berkkanoglu 2010, Berkkanoglu 2017 [pers comm]) (Characteristics of excluded studies).

A further six studies are ongoing (Characteristics of ongoing studies), and one trial is awaiting assessment (Characteristics of studies awaiting classification).

## Risk of bias in included studies

We assessed the risk of bias for each included trial (Characteristics of included studies). We present the results in the 'Risk of bias' summary and graph (Figure 2; Figure 3).

**Figure 2. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.**

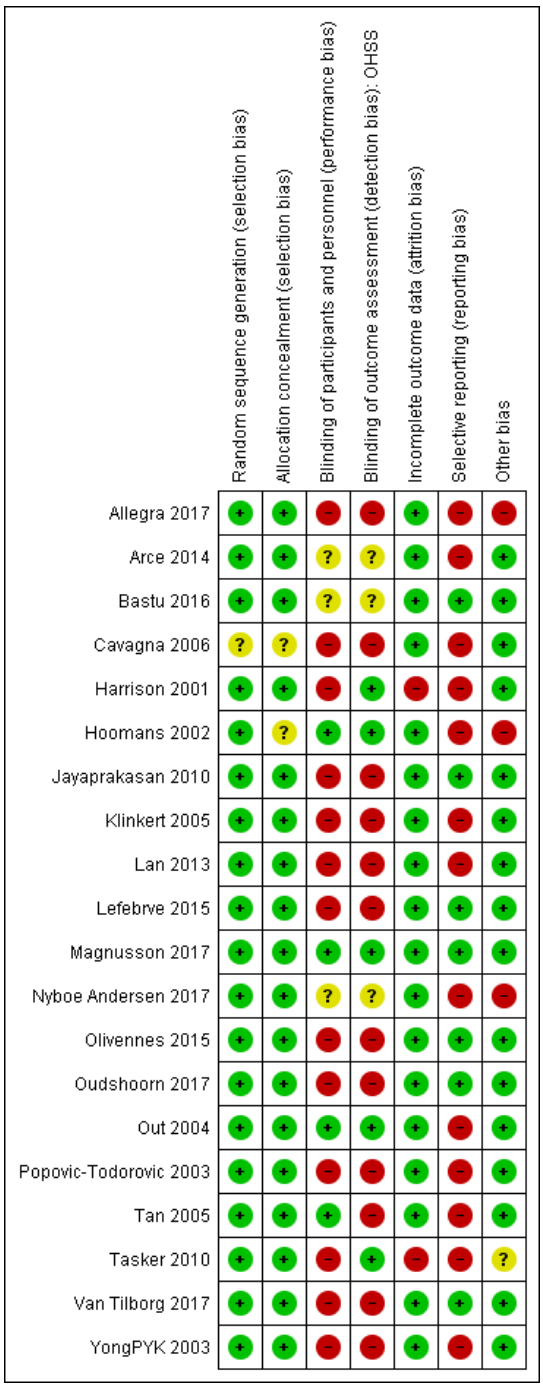
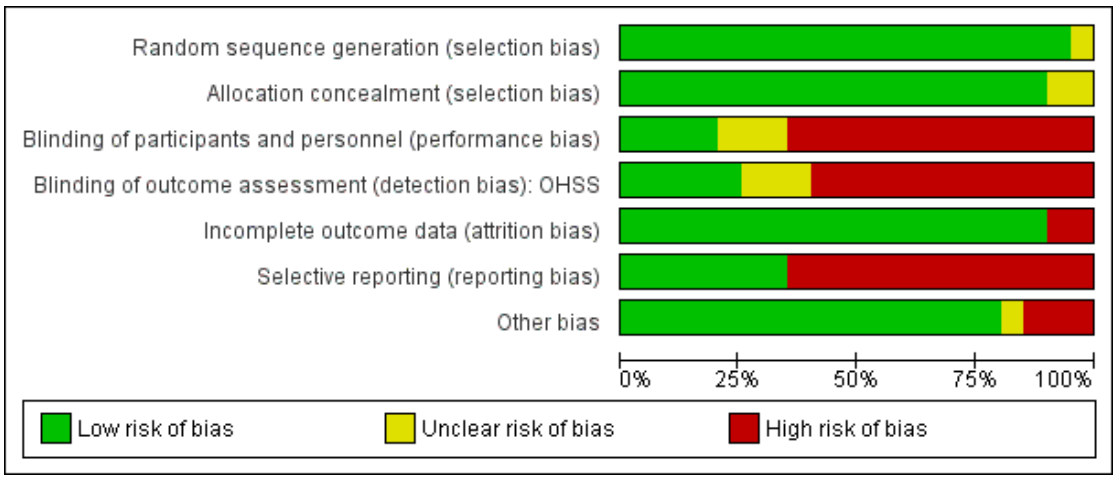


**Figure 3. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.**




### Allocation

Nineteen studies were at low risk of selection bias related to sequence generation, as the studies used computer-generated random numbers. Trialists of one study described it as 'randomized' only, and it was not possible to contact the study authors for further information, therefore we rated the risk of bias for this study as unclear (Cavagna 2006).

Eighteen studies were at low risk of bias allocation concealment, as the studies used SNOSE (Bastu 2016; Klinkert 2005; Lefebrve 2015; Tasker 2010; YongPYK 2003), employed a double-blind design with patient numbers corresponding to boxes containing medication (Out 2004; Tan 2005), concealed allocation within an electronic randomization and case-report system (Arce 2014; Jayaprakasan 2010; Magnusson 2017; Nyboe Andersen 2017; Olivennes 2015; Oudshoorn 2017; Van Tilborg 2017), or used third-party randomization (Allegra 2017; Harrison 2001; Lan 2013; Popovic-Todorovic 2003). We graded two studies as being at unclear risk, as we could not obtain any description of allocation concealment through author correspondence (Cavagna 2006; Hoomans 2002).

### Blinding

### Performance and detection bias

We considered blinding of participants and personnel to be important in this review, as knowledge of trial allocation may impact on the decisions made by staff during the participants' IVF cycle, for example whether to increase or decrease the dose in studies permitting dose adjustment, when to trigger, whether to cancel the cycle for poor or hyper-response, what efforts to make to obtain eggs at egg retrieval, etc. We assessed the domain of detection bias for subjective outcomes only, i.e. OHSS. Indeed, one of the included studies in predicted low responders found that clinicians were more likely to cancel cycles in the lower-dose arm, despite strict rules for cancellation. These authors hypothesised that the treating clinicians were more likely to cancel the cycle in women they knew were on a lower rather than higher dose of FSH.

We judged 11 studies to be at high risk of bias for both domains, as there was no effort made to blind participants, personnel or outcome assessors (Allegra 2017; Cavagna 2006; Jayaprakasan 2010; Klinkert 2005; Lan 2013; Lefebrve 2015; Olivennes 2015; Oudshoorn 2017; Popovic-Todorovic 2003; Van Tilborg 2017; YongPYK 2003). Two studies did not report the only subjective outcome of this study (OHSS), so we rated these as being at low risk, as the domain does not apply (Harrison 2001; Tasker 2010). Six studies employed some level of blinding: in three studies, medications were indistinguishable, and all participants and personnel were blind, so we rated these trials as being at low risk of bias (Hoomans 2002; Magnusson 2017; Out 2004). In three studies, only trial staff were blinded, with no participant blinding (Arce

2014; Bastu 2016; Nyboe Andersen 2017). The studies did not include any description of any safeguards to prevent participants from disclosing their study dose to trial staff, so we rated these studies as being at unclear risk of bias. In a third case, authors described the study as being double-blind; however, the methods seem to indicate that blinding was broken as early as day 4 of FSH administration, which would therefore leave the study open for the most part, warranting a rating of high risk (Tan 2005).

### Incomplete outcome data

We rated 18 studies as being at low risk for incomplete outcome data, as there were few withdrawals or dropouts. Many studies had a number of women who did not reach the stage of embryo transfer and therefore did not have the opportunity to conceive during the study period. We did not consider these participants to contribute to the attrition numbers but rather as not achieving pregnancy or live birth. One study described the exclusion of 19 participants; however, it was not clear which trial arms these participants were excluded from, so it was not possible to assess if the number and reasons were balanced (Harrison 2001). Another study was published as an abstract only, and authors provided the individual participant data from the trial (Hamoda 2017 [pers comm]). The data provided appeared to have a large amount of missing data, and outcomes were not available for a significant number of participants (Tasker 2010). We rated these two studies as being at high risk of bias.

### Selective reporting

A number of studies were at high risk of reporting bias, as they were not registered prospectively and failed to report important outcomes such as live birth and OHSS (Allegra 2017; Cavagna 2006; Harrison 2001; Hoomans 2002; Klinkert 2005; Out 2004; Popovic-Todorovic 2003; Tan 2005; Tasker 2010). Although trial registration was not introduced as mandatory until 2005, the potential for selective reporting remains.

Two studies were at high risk because they either changed the definition of at least one outcome from that listed on the original trial registration (the definition of a good oocyte yield in Allegra 2017) or did not report the same outcomes as those listed (total doses administered in Arce 2014). Another study provided the outcomes of poor response and hyper-response to stimulation only within subgroups of women, and it was not possible to extract the overall data per trial arm (Nyboe Andersen 2017). These authors declined to provide the data per trial arm without providing an adequate reason (Helmgaard 2017b [pers comm]).

Six studies were registered prospectively and reported all outcomes listed at trial registration (Bastu 2016; Jayaprakasan 2010; Magnusson 2017; Olivennes 2015; Oudshoorn 2017; Van Tilborg 2017). Another study listed a number of outcomes on the trial registration that they did not report in the paper; however, the authors provided the data for these outcomes (Lefebrve 2015, Lefebvre 2017 [pers comm]). We rated these five studies as being at low risk of bias.

### Other potential sources of bias

Most studies had no additional sources of possible bias. One study stopped early on the basis the O'Brien and Fleming 1979 rules (O'Brien 1979), which are known to be associated with a biased estimate of effect (Allegra 2017). The analyses in the trial correctly adjusted for the early stopping - however, from our point of view as systematic review authors, the uncorrected summary data available will represent a biased estimate of the treatment effect. One study does not appear to have performed a power calculation, and the decision to complete recruitment on the basis of interim results may have induced bias (Hoomans 2002). Another study was available as an abstract only, therefore detailed information about the study methodology was not available. Although the study authors provided the individual participant data, there were a lot of missing values (Hamoda 2017 [pers comm]; Tasker 2010). We attempted to minimise this bias by performing multiple imputation on the data set, however. One trial performed an interim analysis, and used these interim results to inform a decision to increase the trial sample size; however, there does not appear to be any correction for P value spending (Nyboe Andersen 2017). It is unclear whether or not this would bias the data available for this review.

## Effects of interventions

See: **Summary of findings for the main comparison** ORT-based algorithm compared to standard dose of FSH for women undergoing IVF/ICSI; **Summary of findings 2** Anticipated low-responders: higher compared to lower dose of FSH for women undergoing IVF/ICSI; **Summary of findings 3** Anticipated normal-responders: higher compared to lower dose of FSH for women undergoing IVF/ICSI; **Summary of findings 4** Anticipated high-responders: higher compared to lower dose of FSH for women undergoing IVF/ICSI

We present the results separately for direct dose comparison and ORT-algorithm studies. Within the direct dose comparison studies, we subdivide the results according to each predicted responder category (low, normal, high).

### Direct dose comparison studies

#### 1. Predicted low responders

Five studies included women who were predicted to have a low response based on at least one ORT measure (Bastu 2016; Harrison 2001; Klinkert 2005; Lefebrve 2015; Van Tilborg 2017). We pooled the studies within this comparison in cases where the control dose was identical.

- 300/450 IU versus 150 IU (Klinkert 2005; Van Tilborg 2017)
- 400/450 IU versus 300 IU (Bastu 2016; Harrison 2001)
- 600 IU versus 450 IU (Lefebrve 2015).

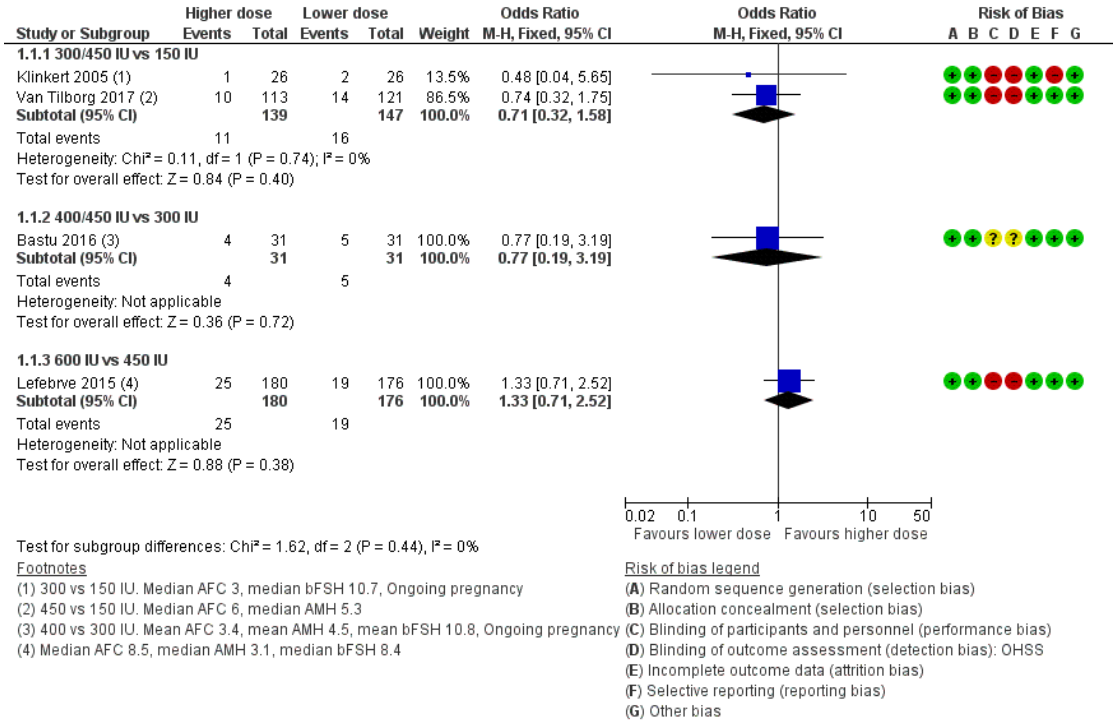
These comparisons are displayed within subgroups on one forest plot for illustrative purposes only (no overall pooling; Summary of findings 2).

**Primary outcomes**

***1.1 Live birth or ongoing pregnancy***

Two studies reported live birth (Lefebrve 2015; Van Tilborg 2017), and two reported ongoing pregnancy (Bastu 2016; Klinkert 2005). The estimates of difference in live birth/ongoing pregnancy rate between the dose-comparisons were very imprecise, and there is little information about the true treatment effect, so we graded the body of evidence as low quality (Analysis 1.1, Figure 4).

**Figure 4. Forest plot of comparison: 1 Anticipated low responders: higher vs lower dose, outcome: 1.1 Live birth or ongoing pregnancy per woman randomised.**




- **300/450 IU versus 150 IU** (OR 0.71, 95% CI 0.32 to 1.58; N = 286; 2 studies; $I^2$ = 0%). This suggests that if the chance of live birth with 150 IU is 11%, then the chance with 300/340 IU would be 3.8% to 16%.
- **400/450 IU versus 300 IU** (OR 0.77, 95% CI 0.19 to 3.19; N = 62; 1 study). This suggests that if the chance of live birth with 300 IU is 16%, then the chance with 400/450 IU would be 3.5% to 38%.
- **600 IU versus 450 IU** (OR 1.33, 95% CI 0.71 to 2.52; N = 356; 1 study). This suggests that if the chance of live birth with 450 IU is 11%, then the chance with 600 IU would be 7.9% to 23%.

One study also reported the outcome of cumulative live birth in two ways (Van Tilborg 2017).

- **Cumulative live birth - following one IVF cycle (fresh and frozen transfers)**. The evidence in relation to cumulative live birth when comparing 450 IU versus 150 IU was also consistent with notable effects in either direction or of no difference (OR 0.78, 95% CI 0.35 to 1.73; N = 234; Analysis 1.18). This suggests that if the chance of cumulative live birth with 150 IU is 13%, then the chance with 450 IU would be 5.1% to 21%.
- **Cumulative live birth - following 18 months of IVF (defined as an ongoing pregnancy leading to a live birth occurring within 18 months of randomization)**. The evidence when comparing 450 IU with 150 IU for cumulative live birth after 18 months of IVF was also consistent with notable effects in either direction or of no difference (OR 0.78, 95% CI 0.46 to 1.32; N = 234; Analysis 1.19). This suggests that if the chance of cumulative live birth with 150 IU is 42%, then the chance with 450 IU would be 25% to 49%.

### *1.2 Severe ovarian hyperstimulation syndrome (OHSS)*

Four studies reported severe OHSS; however, there were no incidents of severe OHSS in any of the studies (Bastu 2016; Klinkert 2005; Lefebrve 2015; Van Tilborg 2017 Analysis 1.2).

## Secondary outcomes

### *1.3 Clinical pregnancy*

All five studies reported this outcome, and we graded the body of evidence as low quality (Analysis 1.3).
One subgroup showed higher pregnancy rates in participants in the lower dosing arm; however, the effect remains imprecise.

- **300/450 IU versus 150 IU** (OR 0.50, 95% CI 0.25 to 1.00; N = 286; 2 studies; $I^2$ = 0%). This suggests that if the chance of clinical pregnancy with 150 IU is 18%, then the chance with 300/450 IU would be between 5.3% and 18%.

In the other two subgroups, the results are imprecise and remain consistent with effects in either direction, or of no effect.

- **400/450 IU versus 300 IU** (OR 0.84, 95% CI 0.26 to 2.69; N = 110; 2 studies; $I^2$ = 0%). This suggests that if the chance of clinical pregnancy with 300 IU is 13%, then the chance with 400/450 IU would be between 3.7% and 28%.
- **600 IU versus 450 IU** (OR 1.14, 95% CI 0.66 to 1.99; N = 356, 1 study). This suggests that if the chance of clinical pregnancy with 450 IU is 16%, then the chance with 600 IU would be 11% to 27%. This study also reported the cumulative clinical pregnancy rate; however, this was not an outcome of this review.

### *1.4 Time to clinical pregnancy*

None of the studies in predicted low responders reported this outcome; Van Tilborg 2017 reported the time to ongoing pregnancy, but not time to clinical pregnancy.

### *1.5 Moderate or severe OHSS*

All five studies reported this outcome; however, there were no incidents of moderate or severe OHSS in three studies (Bastu 2016; Klinkert 2005; Van Tilborg 2017 Analysis 1.5). In the study comparing 600 IU and 450 IU, there was only one occurrence of moderate OHSS in the higher dose arm, so the effect is too imprecise to provide any useful information (Lefebrve 2015).

### *1.6 Multiple pregnancy rate per woman randomized*

Four studies reported multiple pregnancy (Analysis 1.6). In two studies there were no cases of multiple pregnancy in either of the study arms (Bastu 2016; Klinkert 2005). In another there were two events in the 150 IU arm and one in the 450 IU arm (Peto OR 0.55, 95% CI 0.06 to 5.31; N = 234; Van Tilborg 2017). As the event rates were so low, we did not interpret these results any further. In the fourth study there were four multiple pregnancies in the 450 IU arm and eight in the 600 IU arm (Peto OR 0.49, 95% CI 0.16 to 1.55; N = 356; Lefebrve 2015).
We also calculated the multiple pregnancy rates per clinical pregnancy, with similar results (Analysis 1.20).

### *1.7 Number of oocytes retrieved per woman randomized*

All five studies reported this outcome (Analysis 1.7). These are mean differences on the log scale and should not be misinterpreted as numbers of eggs.
In comparing 300/450 IU versus 150 IU, the pooled effect suggests a higher number of oocytes are collected in the higher dose arms (log(MD) oocytes 0.69, 95% CI 0.50 to 0.88; N = 286; 2 studies; $I^2$ = 90%). This pooled estimate should be treated with caution owing to the high statistical heterogeneity.
In the other two comparisons, there did not appear to be any difference in the number of eggs collected; however, the effects remain imprecise, and there could be small effects in either direction.

- **400/450 IU versus 300 IU** (log(MD) oocytes −0.03, 95% CI −0.30 to 0.24; N = 110; 2 studies, $I^2$ = 38%).
- **600 IU versus 450 IU** (log(MD) oocytes 0.08, 95% CI −0.04 to 0.20; N = 356).

### *1.8 Poor response to stimulation*

Two trials within the same subgroup reported this outcome (Analysis 1.8). One study defined a poor response as the collection of fewer than four oocytes or cycle cancellation due to poor response (Klinkert 2005), and the second study defined a poor response as cycle cancellation for poor response or the retrieval of fewer than five oocytes (Van Tilborg 2017). The pooled effect demonstrates that there were fewer cases of poor response among women receiving 300/450 IU than 150 IU.

- **300/450 IU versus 150 IU** (OR 0.52, 95% CI 0.32 to 0.84; N = 286; 2 studies; $I^2$ = 0%). This suggests that if the chance of a poor response with 150 IU is 65%, then the chance with 400/450 IU would be 37% to 61%.

### *1.9 Normal response to stimulation*

None of the studies in predicted low responders reported this outcome specifically in the paper. However, we calculated it as the difference between the number of women randomized and the number with either poor or hyper-response in one trial that reported both of these outcomes (Van Tilborg 2017). Therefore, the resulting definition is women with the retrieval of 5 to 15 oocytes or cycle cancellation for any reason other than a poor or hyper-response (Analysis 1.9). The result suggests there is a higher rate of normal response among women administered 450 IU compared to 150 IU.

- **300/450 IU versus 150 IU** (OR 1.79, 95% CI 1.05 to 3.04; N = 234). This suggests that if the chance of a normal response with 150 IU is 33%, then the chance with 450 IU would be 34% to 60%.

### *1.10 Hyper-response to stimulation*

One study reported this outcome, defining it as cycle cancellation owing to excessive response or more than 15 oocytes at retrieval (Analysis 1.10). The result suggests 450 IU leads to more cases of hyper-response than 150 IU.

- **300/450 IU versus 150 IU** (OR 4.53, 95% CI 0.94 to 21.82; N = 234; Van Tilborg 2017). This suggests that if the chance of a hyper-response with 150 IU is 1.7%, then the chance with 450 IU would be 1.6% to 27%.

### *1.11 Cycle cancellations for poor response*

All five studies reported this outcome (Analysis 1.11). In the first subgroup, the rate of cycle cancellation for poor response was higher among women in the lower-dose group; however, this result is largely influenced by one trial, and heterogeneity remains high.

- **300/450 IU versus 150 IU** (OR 0.23, 95% CI 0.11 to 0.47; N = 286; 2 studies; $I^2$ = 88%). This suggests that if the chance of cycle cancellation for poor response with 150 IU is 28%, then the chance with 300/450 IU would be 4.1% to 15%.

In the other two groupings, the effects of different doses were unclear, and the confidence intervals remain wide.

- **400/450 IU versus 150 IU** (OR 1.47, 95% CI 0.62 to 3.49; N = 110; 2 studies; $I^2$ = 0%). This suggests that if the chance of cycle cancellation for poor response with 150 IU is 22%, then the chance with 400/450 IU would be 15% to 49%.
- **600 IU versus 450 IU** (OR 0.86, 95% CI 0.50 to 1.50; N = 356; 1 study). This suggests that if the chance of cycle cancellation for poor response with 450 IU is 18%, then the chance with 600 IU would be 10% to 25%.

### *1.12 Cycle cancellations for hyper-response or freeze-all*

All studies reported this outcome; however, in four studies there were no events (Bastu 2016; Harrison 2001; Klinkert 2005; Lefebrve 2015). In one study, there was only one event of cycle cancellations for hyper-response, so the effect estimates remain very imprecise (Analysis 1.12).

- **400/450 IU versus 150 IU** (Peto OR 7.93, 95% CI 0.16 to 400.62; N = 234; Van Tilborg 2017).

### *1.13 Cycle cancellations for poor or hyper-response*

This outcome refers to the cancellation of an IVF/ICSI treatment cycle due to poor response or hyper-response, excluding cancellations for other reasons (such as uterine anomaly). In predicted low responders, there was only one case of cancellation for hyper-response, so the outcome primarily reflects the outcome of cycle cancellation for poor response (Analysis 1.13).
In one case there was a clear benefit from 300/450 IU over 150 IU in reducing the number of cycle cancellations from poor or hyper-response.

- **300/450 IU versus 150 IU** (OR 0.25, 95% CI 0.13 to 0.50; N = 286; 2 studies; $I^2$ = 87%). This suggests that if the chance of cycle cancellation for poor or hyper-response with 150 IU is 28%, then the chance with 300/450 IU would be 4.8% to 16%. Readers should treat this estimate with caution owing to the high statistical heterogeneity.

However, in the other two subgroups the effect is less clear.

- **400/450 IU versus 300 IU** (OR 1.47, 95% CI 0.62 to 3.49; N = 110; 2 studies; $I^2$ = 0%). This suggests that if the chance of cycle cancellation for poor or hyper-response with 300 IU is 22%, then the chance with 400/450 IU would be 15% to 49%.
- **600 IU versus 450 IU** (OR 0.86, 95% CI 0.50 to 1.50; N = 356; 1 study). This suggests that if the chance of cycle cancellation for poor or hyper-response with 450 IU is 18%, then the chance with 600 IU would be 10% to 25%.

### *1.14 Proportion of women with at least one transferable embryo*

This outcome refers to the number of women who had at least one embryo available for transfer, either for a fresh embryo transfer or for a freeze-all strategy. In most cases, the estimate of the difference in the number of women with at least one embryo available to transfer between the two groups was very imprecise (Analysis 1.14).

- **400/450 IU versus 300 IU** (OR 0.71, 95% CI 0.31 to 1.60; N = 110; 2 studies; $I^2$ = 0%). This suggests that if the chance of having at least one transferable embryo with 300 IU is 73%, then the chance with 400/450 IU would be 45% to 81%.
- **600 IU versus 450 IU** (OR 1.19, 95% CI 0.78 to 1.82; N = 356; 1 study). This suggests that if the chance of having at least one transferable embryo with 450 IU is 57%, then the chance with 600 IU would be 51% to 71%.

However in one subgroup, more women had at least one transferable embryo among those administered a higher dose.

- **300/450 IU versus 150 IU** (OR 1.76, 95% CI 1.07 to 2.87; N = 286; 2 studies; $I^2$ = 76%). This suggests that if the chance of having at least one transferable embryo with 150 IU is 59%, then the chance with 300/450 IU would be 61% to 81%. Readers should treat this result with caution owing to the high statistical heterogeneity.

### *1.15 Total dose of FSH*

All studies reported this outcome. In all groupings, participants in the higher dosing arm received a higher total dose of FSH on average than women in the lower dosing arm (Analysis 1.15).

- **300/450 IU versus 150 IU** (MD IU 2780, 95% CI 2570 to 3000; N = 286; 2 studies; $I^2$ = 98%).
- **400/450 IU versus 300 IU** (MD IU 1110, 95% CI 910 to 1310; N = 110; 2 studies; $I^2$ = 94%).
- **600 IU versus 450 IU** (MD IU 1200, 95% CI 1070 to 1330; N = 356; 1 study).

We would urge caution in relation to this outcome, however, as the effect of censoring due to cancelled cycles cannot be accounted for, and there is high statistical heterogeneity.

### *1.16 Duration of FSH administration*

Four studies reported this outcome (Analysis 1.16). In all cases, the pooled effects suggest that higher doses reduce the duration of FSH administration.

- **300/450 IU versus 150 IU** (MD days −0.70, 95% CI −1.48 to 0.08; N = 234; 1 study).
- **400/450 IU versus 300 IU** (MD -0.67, 95% CI -1.39 to 0.06; participants = 110; studies = 2; $I^2$ = 0%)
- **600 IU versus 450 IU** (MD days −1.00, 95% CI −1.27 to −0.73; N = 356; 1 study).

We would urge caution in relation to this outcome, however, as it is not possible to account for the effect of censoring due to cancelled cycles.

### *1.17 Cost per woman randomized*

None of the studies in predicted low responders reported this outcome; however, the outcome was available pooled across two sub-studies in this review (one poor responder and one normal responder), which were published as one trial (Van Tilborg 2017). The total cost was higher among women administered 450/225 IU compared to those given 150 IU (EUR 6397 versus EUR 5298; MD EUR 1099, 95% CI 562 to 1591).

#### 2. Predicted normal responders

Nine studies included women with predicted normal response, as determined by at least one ORT measure (Arce 2014; Cavagna 2006; Harrison 2001; Hoomans 2002; Jayaprakasan 2010; Out 2004; Tan 2005; Van Tilborg 2017; YongPYK 2003; Summary of findings 3). The studies are pooled under the following comparisons.

- **200 IU versus 100 IU** (Hoomans 2002; Tan 2005).
- **225/200 IU versus 150 IU**.
  - **200 versus 150 IU** (Cavagna 2006; Harrison 2001; Out 2004).
  - **225 versus 150 IU** (Van Tilborg 2017; YongPYK 2003).
- **300 IU versus 225 IU** (Jayaprakasan 2010).

These comparisons are also presented together in one forest plot for display purposes only (no overall pooling).

- **Dose-response effects (no pooling)**. Arce 2014 reported doses as 5.2 µg, 6.9 µg, 8.6 µg. 10.3 µg, and 12.1 µg rather than as international units (IU). As these doses cannot be translated into the doses described in the other studies, we present information on the dose response between increasing dose groups in separate forest plots and in descriptions in the text.

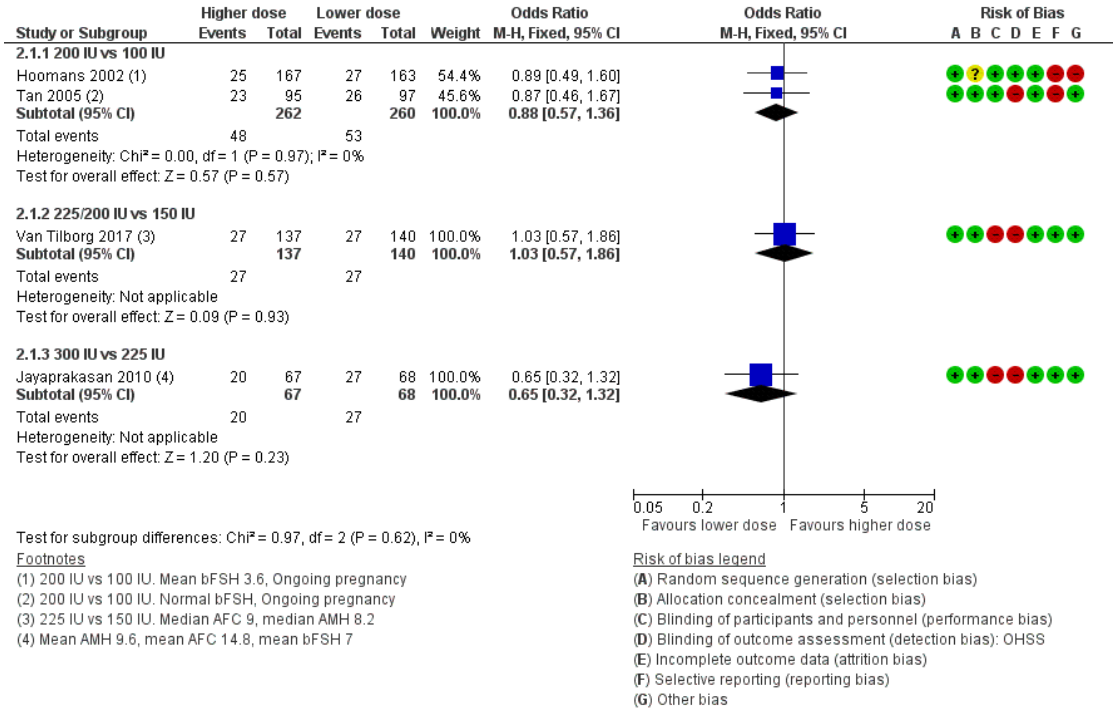
#### Primary outcomes

### *2.1 Live birth or ongoing pregnancy*

Five studies reported this outcome: three reported live birth (Arce 2014; Jayaprakasan 2010; Van Tilborg 2017), and two reported ongoing pregnancy (Hoomans 2002; Tan 2005). We rated the evidence as low quality.

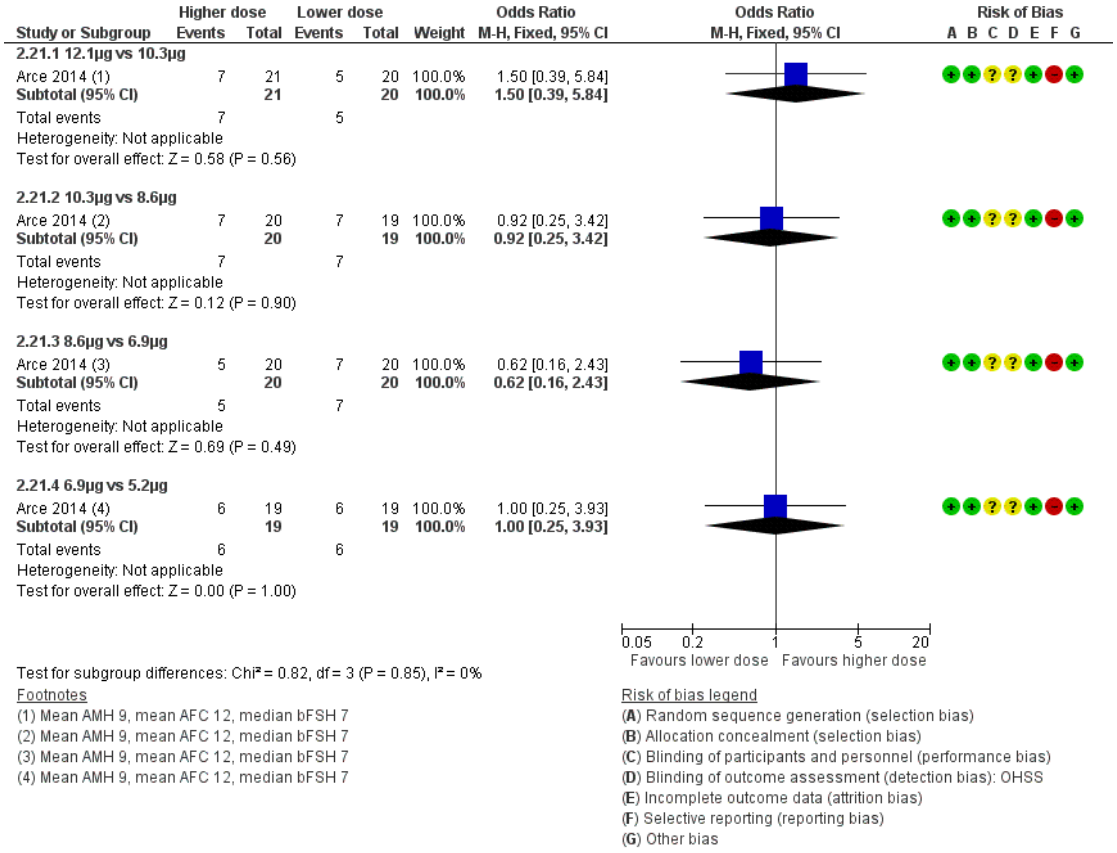
In two of the comparisons, there is no clear impact of different doses on the probability of live birth, and although the confidence intervals encompass the possibility of small effects in either direction, the point estimates sit close to the line of no effect, which makes any benefit from higher doses of FSH unlikely (Analysis 2.1; Figure 5).

**Figure 5. Forest plot of comparison: 2 Anticipated normal responders: higher vs lower dose, outcome: 2.1 Live birth or ongoing pregnancy.**




- **200 versus 100 IU** (OR 0.88, 95% CI 0.57 to 1.36; N = 522; 2 studies; $I^2$ = 0%). This suggests that if the chance of live birth or ongoing pregnancy with 100 IU is 20%, then the chance with 200 IU would be 13% to 26%.
- **225/200 versus 150 IU** (OR 1.03, 95% CI 0.57 to 1.86; N = 277; 1 study). This suggests that if the chance of live birth or ongoing pregnancy with 150 IU is 19%, then the chance with 200/225 IU would be 12% to 31%.
- **300 IU versus 225 IU** (OR 0.65, 95% CI 0.32 to 1.32; N = 135, 1 study). This suggests that if the chance of live birth with 225 IU is 40%, then the chance with 300 IU would be 17% to 47%. In the third comparison, the confidence interval remains wide, and it is not clear whether there is any effect from 300 IU versus 225 IU.
- Additionally, live birth rates and associated standard errors (SEs) across the **five dose groups** in order of increasing dose were 32% (11), 32% (11), 35% (11), 25% (10), and 29% (10). These data neither confirm nor rule out dose effects on live birth (Analysis 2.21; Figure 6).

**Figure 6. Forest plot of comparison: 2 Anticipated normal responders: higher vs lower dose, outcome: 2.21 Dose-response: live birth or ongoing pregnancy.**




Two trials reported cumulative live birth rates, using two definitions (Arce 2014; Van Tilborg 2017).

- **Cumulative live birth - following one IVF cycle (fresh and frozen transfers)** (Analysis 2.18). In the comparison of 200/225 versus 150 IU, the OR was 0.88 (95% CI 0.51 to 1.52; N = 277; 1 study). This suggests that if the chance of cumulative live birth with 150 IU is 26%, then the chance with 225 IU would be between 15% and 35%. Cumulative live birth rates (SEs) across the five dose groups were 37% (11), 42% (11), 35% (11), 30% (10), and 38% (11) (Analysis 2.22). These data neither confirm nor rule out dose effects on cumulative live birth.
- **Cumulative live birth - following 18 months of IVF (defined as an ongoing pregnancy leading to a live birth occurring within 18 months of randomization)**. The evidence in relation to cumulative live birth rate after 18 months of IVF, when comparing 225 IU with 150 IU, was consistent with notable effects in either direction or of no difference (OR 1.01, 95% CI 0.63 to 1.62; N = 277; Analysis 2.19, Van Tilborg 2017). This suggests that if the chance of cumulative live birth with 150 IU is 47%, then the chance with 225 IU would be between 36% and 59%.

### *2.2 Severe OHSS*

Eight of the nine trials reported this outcome; however, there were no incidents of severe OHSS in four of the studies (Arce 2014; Cavagna 2006; Hoomans 2002; YongPYK 2003). We graded the body of evidence as of very low quality, and the effect estimates remain very imprecise.

- **200 IU versus 100 IU** (Peto OR 0.14, 95% CI 0.00 to 6.96; N = 522; 2 studies; $I^2$ = 0%; Analysis 2.2).
- **225/200 IU versus 150 IU** (Peto OR 1.00, 95% CI 0.20 to 5.02; N = 740; 4 studies; $I^2$ = 15%; Analysis 2.2).

- **300 IU versus 225 IU** (Peto OR 0.14, 95% CI 0.00 to 6.92; N = 135, 1 study; Analysis 2.2).
- In the **multiple-dosing trial** there were no cases of severe OHSS (Arce 2014).

**Secondary outcomes**

### *2.3 Clinical pregnancy*

Eight studies in predicted normal responders reported this outcome (Analysis 2.3). In each case the point estimates suggest no benefit from increased doses on the probability of pregnancy; however, the estimates are imprecise and consistent with small effects in either direction. We graded the body of evidence as low quality.

- **200 IU versus 100 IU** (OR 0.86, 95% CI 0.50 to 1.49; N = 330; 1 study). This suggests that if the chance of clinical pregnancy with 100 IU is 20%, then the chance with 200 IU would be 11% to 27%.
- **225/200 IU versus 150 IU** (OR 0.98, 95% CI 0.73 to 1.31; N = 1037; 5 studies; $I^2$ = 0%). This suggests that if the chance of clinical pregnancy with 150 IU is 24%, then the chance with 200/225 IU would be 18% to 29%.
- **300 IU versus 225 IU** (OR 0.91, 95% CI 0.46 to 1.80; N = 135, 1 study). This suggests that if the chance of clinical pregnancy with 225 IU is 44%, then the chance with 300 IU would be 27% to 59%.
- In the **multiple-dosing study**, clinical pregnancy rates were essentially identical to birth rates, since only one pregnancy did not progress to live birth; rates (SEs) across the five dose groups were 31% (11), 32% (11), 35% (11), 25% (10), and 33% (10) (Analysis 2.23). These data neither confirm nor rule out dose effects on clinical pregnancy.

### *2.4 Time to clinical pregnancy*

None of the studies in predicted normal responders reported this outcome (Van Tilborg 2017 reported the time to ongoing pregnancy, but not time to clinical pregnancy).

### *2.5 Moderate or severe OHSS*

The estimates for this outcome are based on a small number of events in eight studies, and therefore the effect estimates remain imprecise, and we rated the evidence as very low quality (Analysis 2.5).

- **200 IU versus 100 IU** (Peto OR 0.62, 95% CI 0.21 to 1.87; N = 522; 2 studies; $I^2$ = 0%). This suggests that if the chance of moderate or severe OHSS with 100 IU is 3.1%, then the chance with 200 IU would be 0.7% to 5.6%.
- **225/200 IU versus 150 IU** (Peto OR 1.21, 95% CI 0.51 to 2.85; N = 740; 4 studies; $I^2$ = 49%) This suggests that if the chance of moderate or severe OHSS with 150 IU is 2.7%, then the chance with 225/200 IU would be 1.4% to 7.3%.
- **300 IU versus 225 IU** (Peto OR 0.67, 95% CI 0.11 to 3.99; N = 135, 1 study). With only five total events, we refrain from interpreting this result any further.
- There were no incidents of moderate or severe OHSS observed in any of the study arms of the **multiple-dosing study** (Arce 2014).

### *2.6 Multiple pregnancy in randomized women*

Five trials reported the outcome of multiple pregnancy, and as there were only a small number of events in each comparison, the point estimates were imprecise and consistent with substantial effects in either direction (Analysis 2.6).

- **200 IU versus 100 IU** (Peto OR 0.97, 95% CI 0.38 to 2.52; N = 330; 1 study). This suggests that if the chance of multiple pregnancy with 100 IU is 5.5% then the chance with 200 IU would be between 2.2% and 13%.
- **225/200 IU versus 150 IU** (Peto OR 1.91, 95% CI 0.38 to 9.69; N = 400; 2 studies; $I^2$ = 0%).
- **300 IU versus 225 IU** (Peto OR 7.61, 95% CI 0.47 to 123.02; N = 135; Jayaprakasan 2010). In this study there were only two multiple pregnancies, both in the 300 IU arm.
- There were no multiple pregnancies in the **multiple-dosing study** (Arce 2014).

We also analyzed the data as multiple pregnancy in women with clinical pregnancy, with similar results (Analysis 2.20).

### *2.7 Number of oocytes per woman randomized*

These are mean differences on the log scale, and should not be misinterpreted as numbers of eggs.

All studies in normal responders reported this outcome (Analysis 2.7). The first two comparisons suggest a higher egg yield from higher doses of FSH.

- **200 IU versus 100 IU** (log(MD) oocytes 0.46, 95% CI 0.36 to 0.57; N = 330; 2 studies, $I^2$ = 98%). Readers should interpret this pooled estimate with caution owing to the high observed statistical heterogeneity.
- **225/200 IU versus 150 IU** (log(MD) oocytes 0.16, 95% CI 0.08 to 0.24; N = 463; 5 studies, $I^2$ = 44%).
- **300 IU versus 225 IU** (log(MD) oocytes 0.03, 95% CI −0.17 to 0.23, N = 135, 1 study). In the third comparison, the evidence did not rule out differences in either direction in the number of oocytes retrieved depending on FSH dose.
- In the **multiple-dosing study**, the mean (standard deviation; SD) numbers of oocytes collected across dose groups

were 4 (2.5), 6 (5.1), 7 (4), 6.9 (3.8), and 9 (5.1) (Analysis 2.24); these have been recalculated to include cancelled cycles, and SDs estimated according to the method described in Data synthesis. We also note that the authors reported a dose-response effect, although their analysis excluded small numbers of cancelled cycles.

### *2.8 Poor response to stimulation*

Two trials reported poor response to stimulation, defining it either as obtaining no more than three oocytes (Arce 2014), or cycle cancellation owing to insufficient growth/five oocytes or fewer at retrieval (Van Tilborg 2017). In both comparisons there were fewer cases of poor response among women administered higher doses of FSH.

- **225 IU versus 150 IU** (OR 0.50, 95% CI 0.30 to 0.83; N = 277; 1 study; Analysis 2.8). This suggests that if the chance of a poor response with 150 IU is 40% then the chance with 225 IU would be between 17% and 36%.
- In the **multiple-dosing study** the proportion (SE) of participants with poor response across the dose groups was 37% (11), 32% (11), 20% (9), 10% (7), and 14% (8) (Analysis 2.25). These data neither confirm nor rule out dose effects on poor response.

### *2.9 Normal response to stimulation*

Only two trials reported this outcome. Arce 2014 defined a normal response as obtaining 4 to 14 oocytes. Van Tilborg 2017 calculated the outcome as the difference between the number of women randomized and the number with either poor or hyper-responses. Therefore the resulting definition is women in whom 5 to 15 oocytes were retrieved or whose cycle was cancelled for any reason other than a poor or hyper-response. There was no clear difference in the occurrence of normal response in different dose comparisons.

- **225 IU versus 150 IU** (OR 1.26, 95% CI 0.78 to 2.04; N = 277; 1 study; Analysis 2.9). This suggests that if the chance of a normal response with 150 IU is 56% then the chance with 225 IU would be between 50% and 73%.
- In the **multiple dosing study** the proportion (SE) of participants with normal response across the dose groups was 63% (11), 58% (11), 75% (10), 90% (7), and 76% (9) (Analysis 2.26).

### *2.10 Hyper-response to stimulation*

The same two studies reported this outcome, with Arce 2014 defining it as women with 15 or more eggs collected and Van Tilborg 2017 as women with either more than 15 eggs collected or cancellation for hyper-response. In one trial there were more cases of hyper-response among women in the higher dosing arm.

- **225 IU versus 150 IU** (OR 4.08, 95% CI 1.47 to 11.34; N = 277; 1 study; Analysis 2.10). This suggests that if the chance of a hyper-response with 150 IU is 3.6% then the chance with 225 IU would be between 5.2% and 30%.
- In the **multiple-dosing study** the proportion (SE) of participants with hyper-response across the dose groups was 0% (0), 11% (7), 5% (5), 0% (0), and 10% (6) (Analysis 2.27). These data neither confirm nor rule out dose effects on hyper-response.

### *2.11 Cycle cancellations for poor response*

This outcome was available for all studies. In all cases, administration of a higher dose of FSH led to fewer cycle cancellations for poor response; however, the effect was of variable magnitude and precision.

- **200 IU versus 100 IU** (OR 0.33, 95% CI 0.16 to 0.66; N = 522; 2 studies; $I^2$ = 60%; Analysis 2.11). This suggests that if the chance of cancellation for poor response with 100 IU is 11%, then the chance with 200 IU would be 4.1% to 9.5%.
- **225/200 IU versus 150 IU** (OR 0.56, 95% CI 0.36 to 0.88; N = 1037; 5 studies; $I^2$ = 13%; Analysis 2.11). This suggests that if the chance of cancellation for poor response with 150 IU is 10%, then the chance with 225/220 IU would be 4.0% to 9.4%.
- **300 IU versus 225 IU.** There was only one cancellation for poor response in both arms (OR 0.11, 95% CI 0.01 to 2.01; N = 135, 1 study; Analysis 2.11).
- Cancellation rates for poor response (SEs) across the dose groups in the **multiple-dosing study** were 11% (7), 5% (5), 5% (5), 0% (0), and 5% (5) (Analysis 2.28). These data neither confirm nor rule out dose effects on cycle cancellation for poor response.

### *2.12 Cycle cancellations for hyper-response or freeze-all per woman randomized*

This outcome refers to cycle cancellations only for reasons of poor or hyper-response (excluding cancellations for other reasons), and all studies reported it. As the occurrence of cancellation for hyper-response in this population was low, the pooled effects are associated with a large degree of imprecision (Analysis 2.12).

- **200 IU versus 100 IU** (Peto OR 1.93, 95% CI 0.20 to 18.62; N = 522; 2 studies; $I^2$ = 62%).
- **225/200 IU versus 150 IU** (Peto OR 2.28, 95% CI 0.99 to 5.26; N = 1037; 5 studies; $I^2$ = 0%). This suggests that if the chance of cancellation for hyper-response with 150 IU is 1.4%, then the chance with 225/200 IU would be 1.3% to 6.7%.

- In the comparison of **300 IU versus 225 IU**, there was only one cancellation for hyper-response in both arms (OR 1.02, 95% CI 0.06 to 16.57; N = 135, 1 study).
- In the **multiple-dosing study** there were no cycles cancelled for hyper-response in any of the five dose arms (Arce 2014).

### *2.13 Cycle cancellations for poor or hyper-response*

In combining the cycle cancellations for hyper and poor response (2.11 and 2.12 above), most comparisons suggest that a higher dose is associated with fewer cancellations overall (Analysis 2.13).

- **200 IU versus 100 IU** (OR 0.37, 95% CI 0.19 to 0.72; N = 522; 2 studies; $I^2$ = 10%). This suggests that if the chance of cycle cancellation with 100 IU is 13%, then the chance with 200 IU would be between 2.7% to 9.5%.
- **225/200 IU versus 150 IU** (OR 0.76, 95% CI 0.51 to 1.13; N = 1037; 5 studies; $I^2$ = 0%). This suggests that if the chance of cycle cancellation with 150 IU is 12%, then the chance with 225/200 IU would be between 6.5% to 13%.
- **300 IU versus 225 IU** (OR 0.19, 95% CI 0.02 to 1.68; N = 135, 1 study). This suggests that if the chance of cycle cancellation with 225 IU is 7.4%, then the chance with 300 IU would be between 0.2% to 15%.
- In the **multiple-dosing study** there were only cancellations for poor response, therefore the rates (SE) for total cancellation are the same as presented above across the dose groups: 11% (7), 5% (5), 5% (5), 0% (0), and 5% (5) (Analysis 2.28). These data neither confirm nor rule out dose effects on cycle cancellation for poor response.

### *2.14 Women with at least one transferable embryo*

This outcome refers to the number of women who had at least one embryo available for transfer, either for a fresh embryo transfer or for a freeze-all strategy. All studies reported or calculated this outcome and suggest possible benefit from higher dose; however, this was not as clear in some comparisons (Analysis 2.14).

- **200 IU versus 100 IU** (OR 1.58, 95% CI 0.95 to 2.64; N = 522; 2 studies; $I^2$ = 78%). This suggests that if the chance of having at least one transferable embryo with 100 IU is 84%, then the chance with 200 IU would be 83% to 93%.
- **225/200 IU versus 150 IU** (OR 1.06, 95% CI 0.76 to 1.47; N = 1037; 5 studies; $I^2$ = 64%). This suggests that if the chance of having at least one transferable embryo with 150 IU is 83%, then the chance with 225/200 IU would be 79% to 88%.
- **300 IU versus 225 IU** (OR 1.75, 95% CI 0.60 to 5.13; N = 135, 1 study). This similarly unclear evidence suggests that if the chance of having at least one transferable embryo with 225 IU is 85%, then the chance with 300 IU would be 78% to 97%.
- In the **dose-response study** the percentages (SEs) with at least one embryo available were 89% (7), 84% (8), 90% (7), 85% (8), and 90% (6) (Analysis 2.29). These data neither confirm nor rule out dose effects on women having at least one transferable embryo.

### *2.15 Total dose of FSH*

In all comparisons, participants randomized to a higher daily dose of FSH received a higher total dose during their IVF/ICSI cycle (Analysis 2.15).

- **200 IU versus 100 IU** (MD IU 795.79, 95% CI 656.67 to 934.91; N = 522; 2 studies; $I^2$ = 8%).
- **225/200 IU versus 150 IU** (MD IU 503.12, 95% CI 456.23 to 550.00; N = 1037; 5 studies; $I^2$ = 71%). Readers should treat this pooled estimate with caution owing to the high statistical heterogeneity.
- **300 IU versus 225 IU** (MD IU 725, 95% CI 597 to 853; N = 135, 1 study).
- In the **multiple-dosing study**, mean (SD) total doses were 47.9 µg (12.0), 59.2 µg (12.7), 72.7 µg (11.7), 80.9 µg (15.0), and 95.1 µg (28.7) (Analysis 2.30). It is not possible to convert these to IU to permit pooling with the other trials.

However, we recommend caution when interpreting these results, as cancellations for poor response were essentially censored, making it impossible to impute doses; these cancellations were considerably more common in the lower dosing arms.

### *2.16 Duration of FSH administration*

Eight trials reported this outcome (Analysis 2.16). In most cases, a higher dose of FSH was associated with a shorter duration of FSH; however, we recommend caution when interpreting these results, as cancellations were censored, making it impossible to impute days of stimulation; these cancellations were considerably more common in the lower dosing arms.

- **200 IU versus 100 IU** (MD days −1.80, 95% CI −2.21 to −1.39; N = 330; 1 study).
- **225/200 IU versus 150 IU** (MD days −0.25, 95% CI −0.51 to 0.01; N = 961; 4 studies; $I^2$ = 76%). This pooled estimate should be treated with caution owing to the high statistical heterogeneity.
- **300 IU versus 225 IU** (MD days −0.30, 95% CI −0.79 to 0.19; N = 135, 1 study).
- In the **multiple-dosing study**, mean (SD) days stimulated were 9.2 (2.3), 8.6 (1.8), 8.5 (1.4), 7.9 (1.5), and 7.9 (2.4) (Analysis 2.31), and we note tentatively that the mean duration decreases across the groups.

### *2.17 Cost per woman randomized*

None of the studies in predicted normal responders reported this outcome; however, the outcome was available pooled across two sub-studies in this review (one low responder and one normal responder) which were published as one trial (Van Tilborg 2017). The total cost was higher among women administered 450/225 IU compared to those given 150 IU (EUR 6397 versus EUR 5298; MD EUR 1099, 95% CI 562 to 1591).

#### 3. Predicted high responders

There are two studies included in this comparison, one of which was a dose-response study including 123 predicted high responder participants, reporting outcomes in groups given five doses (5.2 µg, 6.9 µg, 8.6 µg, 10.3 µg, and 12.1µg) of a novel gonadotropin (follitropin delta), so translation to IU is not possible (Arce 2014; Summary of findings 4). The second trial compared doses of 150 IU versus 100 IU in women with AFC over 15 (Oudshoorn 2017). Data from the two studies are presented in separate forest plots, as below.

- **150 IU versus 100 IU** (Oudshoorn 2017).
- **Dose-response effects (no pooling)**: 5.2 µg, 6.9 µg, 8.6 µg. 10.3 µg, 12.1 µg (Arce 2014). These doses cannot be translated into the doses described in the other studies, so we present information on the dose response between increasing dose groups in separate forest plots and describe them in the text.

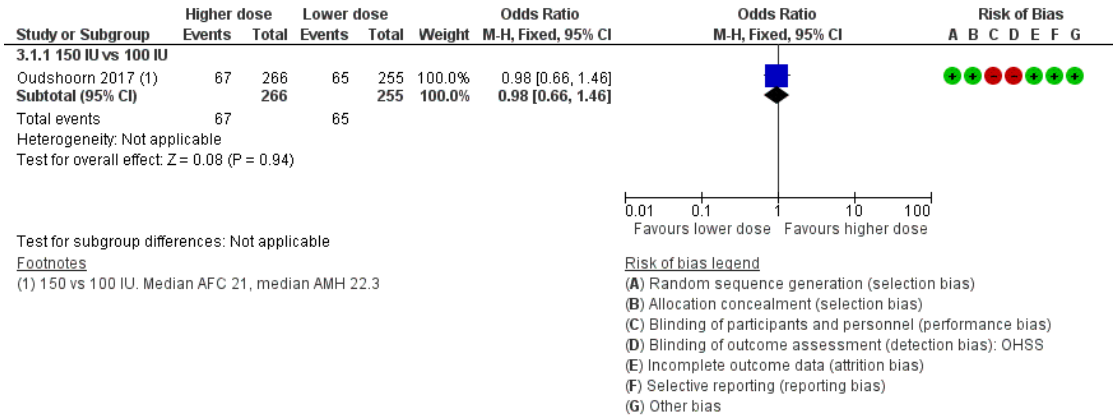
**Primary outcomes**

*3.1 Live birth or ongoing pregnancy*

Both trials reported live birth, and in both cases there was insufficient evidence to determine whether there was any difference in live birth rate between the doses investigated; we cannot rule out moderate effects in either direction. We graded the body of evidence as low quality.

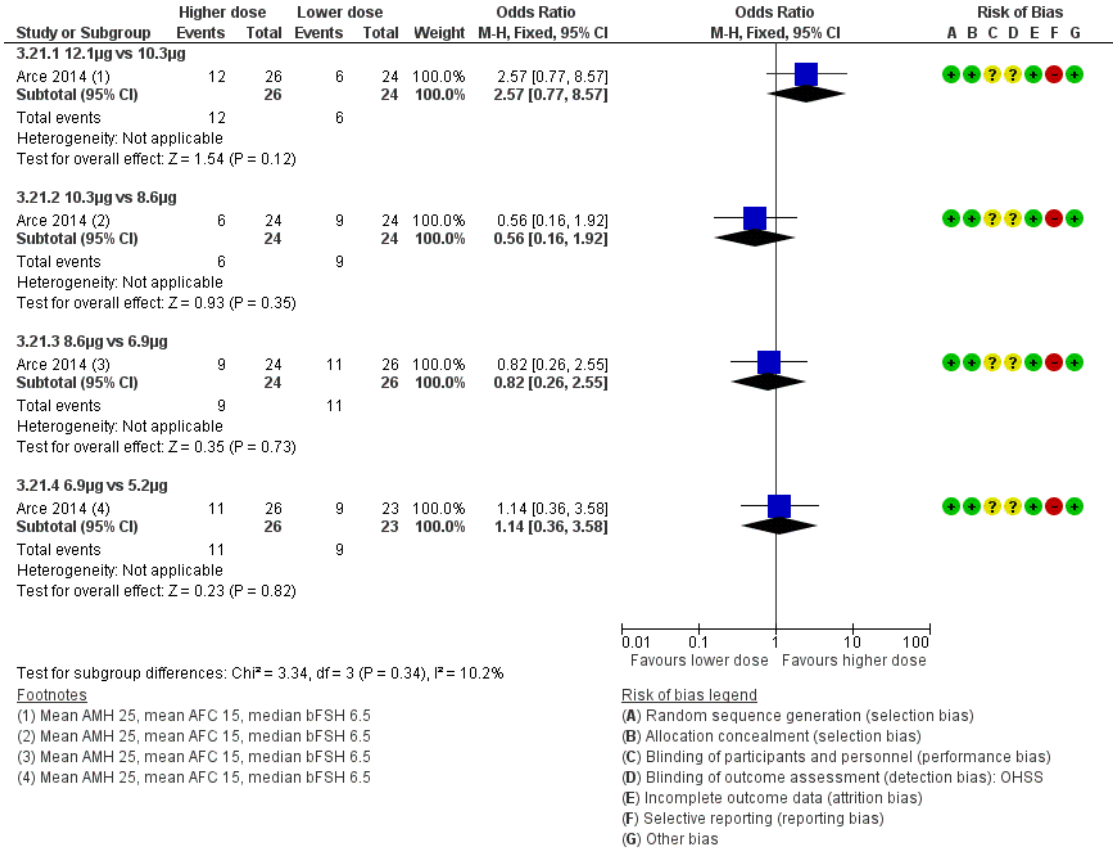
- **150 IU versus 100 IU** (OR 0.98, 95% CI 0.66 to 1.46; N = 521; 1 study; Analysis 3.1; Figure 7). This suggests that if the chance of live birth with 100 IU is 26%, then the chance with 150 IU would be 18% to 33%.

**Figure 7. Forest plot of comparison: 3 Anticipated high-responders: higher vs lower dose, outcome: 3.1 Live birth or ongoing pregnancy.**




- Birth rates (SEs) across the **five dose groups** were 39% (10), 42% (10), 38% (10), 25% (9), and 46% (10) (Analysis 3.21; Figure 8). These data neither confirm nor rule out dose effects on live birth.

**Figure 8. Forest plot of comparison: 3 Anticipated high-responders: higher vs lower dose, outcome: 3.21 Dose-response: live birth or ongoing pregnancy.**




Additionally, both trials reported the cumulative live birth rate, one using two definitions.

- **Cumulative live birth rate - following one IVF cycle (fresh and frozen transfers)**.
  - 150 IU versus 100 IU (OR 1.16, 95% CI 0.81 to 1.65; N = 521; 1 study; Analysis 3.18). This suggests that if the chance of cumulative live birth with 100 IU is 36%, then the chance with 150 IU would be between 31% and 48%.
  - Cumulative birth rates (SEs) across the five dose groups were 43% (10), 54% (10), 46% (10), 38% (10), and 50% (10) (Analysis 3.23). These data neither confirm nor rule out dose effects on live birth.
- **Cumulative live birth - following 18 months of IVF (defined as an ongoing pregnancy leading to a live birth occurring within 18 months of randomization)** (Analysis 3.19).**150 IU versus 100 IU** (OR 1.16, 95% CI 0.80 to 1.68; N = 521; 1 study). This suggests that if the chance of cumulative live birth with 100 IU is 66%, then the chance with 150 IU would be between 61% and 77%.

### *3.2 Severe OHSS*

Both trials reported severe OHSS, and in both cases there were too few events to make any assessment regarding the effect of the different doses on the probability of this outcome. We graded the body of evidence as very low quality.

- 150 IU versus 100 IU (Peto OR 0.72, 95% CI 0.16 to 3.19; N = 521; 1 study; Analysis 3.2).
- Severe OHSS rates (SEs) were 0% (0), 0% (0), 4% (4), 0% (0), and 8% (5) (Analysis 3.24). The event rates are too low to make a reasonable inference regarding a treatment effect.

## Secondary outcomes

### *3.3 Clinical pregnancy*

Both trials reported clinical pregnancy rate, and in both cases there was no obvious benefit of higher or lower dose on the probability of pregnancy; however, the evidence is consistent with moderate effects in either direction. We graded the body of evidence as low quality.

- **150 IU versus 100 IU** (OR 1.14, 95% CI 0.78 to 1.66; N = 521; 1 study; Analysis 3.3). This suggests that if the chance of clinical pregnancy with 100 IU is 28%, then the chance with 150 IU would be 23% to 39%.
- Clinical pregnancy rates (SEs) across the **five dose groups** were 39% (10), 46% (10), 38% (10), 25% (9), and 46% (10) (Analysis 3.22). These data neither confirm nor rule out dose effects on clinical pregnancy.

### *3.4 Time to clinical pregnancy*

None of the studies in predicted high responders reported this outcome (Oudshoorn 2017 reported the time to ongoing pregnancy, but not time to clinical pregnancy).

### *3.5 Moderate or severe OHSS*

Both trials reported moderate/severe OHSS, and in both cases there was a trend towards increased risk of moderate/severe OHSS with increasing dose; however, even a small reduction in risk is possible.

- **150 IU versus 100 IU** (Peto OR 2.31, 95% CI 0.80 to 6.67; N = 521; 1 study; Analysis 3.5). This suggests that if the chance of moderate or severe OHSS with 100 IU is 1.6%, then the chance with 150 IU would be 1.3% to 9.6%.
- Rates (SE) of moderate or severe OHSS across the **five dose groups** were 0% (0), 0% (0), 4% (4), 4% (4), and 12% (6) (Analysis 3.25). Event rates are too low to make any reasonable inference here. We rated the body of evidence for this outcome as very low quality.

### *3.6 Multiple pregnancy in randomized women*

Both trials reported this outcome; however, the number of multiple pregnancies was very small, so we do not interpret the results further.

- **150 IU versus 100 IU** (Peto OR 1.87, 95% CI 0.19 to 18.09; N = 521; 1 study; Analysis 3.6). The multiple pregnancy rates were also calculated per clinical pregnancy, with similar results (Analysis 3.20).
- There were no multiple pregnancies in the **multiple-dosing arm** study (Arce 2014).

### *3.7 Number of oocytes per woman randomized*

Both trials reported this outcome and indicate higher oocyte yield with higher doses of FSH. These are mean differences are on the log scale and should not be misinterpreted as numbers of eggs.

- **150 IU versus 100 IU** (log(MD) oocytes 0.67, 95% CI 0.55 to 0.79; N = 521; 1 study; Analysis 3.7).
- The mean (SD) numbers of eggs collected across the **five dose groups** in the second study were 5.9 (3.9), 9.1 (6.4), 10.2 (5), 13.6 (7.8), and 14.4 (5.8) (Analysis 3.26), where these have been recalculated to include cancelled cycles, and SDs have been estimated according to the method described in Data synthesis. We note tentatively that the means increase with increasing dose. We also note that the authors reported a dose-response effect, although their analysis excluded small numbers of cancelled cycles.

### *3.8 Poor response to stimulation*

One study defined a poor response as cycle cancellation for poor response or the retrieval of fewer than five oocytes (Oudshoorn 2017), and the second study defined poor response as the retrieval of fewer than four oocytes (Arce 2014).

- **150 IU versus 100 IU** (OR 0.15, 95% CI 0.09 to 0.25; N = 521; 1 study; Analysis 3.8). This suggests that if the chance of a poor response with 100 IU is 36%, then the chance with 150 IU would be 4.8% to 12%.
- Rates (SE) of poor response across the **five dose groups** in the second trial were 35% (10), 15% (7), 8% (6), 8% (6), and 0% (0) (Analysis 3.27). These data neither confirm nor rule out dose effects on poor response, although we note that the rate decreases across the dose groups.

### *3.9 Normal response to stimulation*

Both trials reported this outcome. One study reported poor and hyper-response and calculated the normal events as the difference between them, the resulting definition being the number of women with 5 to 15 oocytes collected or with cycle cancellation of any reason other than poor or hyper-response (Oudshoorn 2017). The second trial defined a normal response as the retrieval of 4 to 14 oocytes (Arce 2014). The results suggest a similar number of women achieving a normal response among those administered different doses.

- **150 IU versus 100 IU** (OR 1.07, 95% CI 0.76 to 1.50; N = 521; 1 study; Analysis 3.9). This suggests that if the chance of a normal response with 100 IU is 53%, then the chance with 150 IU would be 46% to 62%.
- The proportion of participants (SE) with normal response across the **five dose groups** was 57% (10), 58% (10), 79% (8), 54% (10), and 50% (10) (Analysis 3.28). These data neither confirm nor rule out dose effects on normal response.

### *3.10 Hyper-response to stimulation*

One study defined a hyper-response as cycle cancellation for excessive response or the retrieval of more than 15 oocytes (Oudshoorn 2017), and the second study defined it as collection of more than 14 oocytes (Arce 2014). There were significantly more women having a hyper-response among those administered 150 IU compared to those given 100 IU.

- **150 IU versus 100 IU** (OR 5.04, 95% CI 3.17 to 8.02; N = 521; 1 study; Analysis 3.10). This suggests that if the chance of a hyper-response with 100 IU is 11%, then the chance with 150 IU would be 28% to 50%.
- The proportion of participants (SE) with hyper-response across the **five dose groups** was 9% (6), 19% (8), 21% (8), 38% (10), and 50% (10) (Analysis 3.29). These data neither confirm nor rule out dose effects on hyper-response, although we note that the rate increases across the dose groups.

### *3.11 Cycle cancellations for poor response*

In one trial, the pooled results suggest that cancellation for poor response occurs more in women administered 100 IU than those given 150 IU.

- **150 IU versus 100 IU** (OR 0.13, 95% CI 0.06 to 0.28; N = 521; 1 study; Analysis 3.11). This suggests that if the chance of a cycle cancellation for poor response with 100 IU is 21%, then the chance with 150 IU would be 1.5% to 6.8%.
- Cancellation rates for poor response (SEs) across the **five dose groups** were 0% (0), 0% (0), 4% (5), 0% (0), and 0% (0) (Analysis 3.30). The event rates are too low to allow any meaningful inference.

### *3.12 Cycle cancellations for hyper-response*

The pooled results suggest that cancellation for hyper-response occurs more in women administered 150 IU compared to 100 IU.

- **150 IU versus 100 IU** (OR 5.28, 95% CI 2.16 to 12.90; N = 521; 1 study; Analysis 3.12). This suggests that if the chance of a cycle cancellation for hyper-response with 100 IU is 2.4%, then the chance with 150 IU would be 4.9% to 24%.
- Cancellation rates (SEs) for hyper-response across the **five dose groups** were 0% (0), 4% (4), 0% (0), 4% (4), and 0% (0) (Analysis 3.31).

### *3.13 Cycle cancellations for poor or hyper-response*

The rate of cycle cancellation for either a poor or hyper-response (combining 3.11 and 3.12 above) was higher among women given 100 IU than in women given 150 IU, as this outcome was denominated by higher overall rates of cancellation for poor response than for hyper-response.

- 150 IU versus 100 IU (OR 0.57, 95% CI 0.36 to 0.89; N = 521; 1 study; Analysis 3.13). This suggests that if the chance of a cycle cancellation for poor or hyper-response with 100 IU is 23%, then the chance with 150 IU would be 9.8% to 21%.
- Cancellation rates (SEs) across the five dose groups were 0% (0), 4% (4), 4% (4), 4% (4), and 0% (0) (Analysis 3.32). Event rates are too low to permit inference.

### *3.14 Women with at least one transferable embryo*

This outcome refers to the number of women who had at least one embryo available for transfer, either for a fresh embryo transfer or for a freeze-all strategy.

- **150 IU versus 100 IU** (OR 2.33, 95% CI 1.53 to 3.55; N = 521; 1 study; Analysis 3.14). This suggests that if the chance of a having at least one embryo for transfer with 100 IU is 69%, then the chance with 150 IU would be significantly higher: 77% to 89%.
- The rate (SEs) of women with at least one transferable embryo across the **five dose groups** were 91% (6), 88% (6), 92% (6), 79% (8), and 96 (4) (Analysis 3.33). No dose-response is evident, and the second highest dose group appears to have fewer women with transferable embryos.

### *3.15 Total dose of FSH*

- **150 IU versus 100 IU**. The total dose of FSH administered was higher among women administered a higher daily dose of FSH (MD IU 345.00, 95% CI 280.34 to 409.66; N = 521; Analysis 3.15).
- In the **multiple-dosing study**, mean total doses (SDs) were 51.8 (11.2), 64.0 (14.3), 71.3 (16.1), 81.1 (13.7), and 100.0 (14.7) (Analysis 3.34). It is not possible to carry out any robust analysis using these summary measures, although we note tentatively that the mean total dose decreases across the groups.

We would urge caution in relation to this outcome, however, as it is not possible to account for the effect of censoring due to cancelled cycles.

### *3.16 Duration of FSH administration*

- **150 IU versus 100 IU**. The pooled results demonstrate that the duration of FSH administration was less in women administered the higher dose (MD days −1.40, 95% CI −1.91 to −0.89; N = 521; 1 study; Analysis 3.16).
- In the **multiple-dosing study**, mean days stimulated (SDs) were 10 (2.2), 9.3 (2.1), 8.3 (1.9), 7.9 (1.3), and 8.3 (1.2) (Analysis 3.35). It isn't possible to carry out any robust analysis using these summary measures, although we note tentatively that the mean duration decreases across the first four groups.

We would urge caution in relation to this outcome however, as it is not possible to account for the effect of censoring due to cancelled cycles.

### *3.17 Cost per woman randomized*

None of the studies in predicted high responders reported this outcome according to the review outcomes; however, the outcome was available for Oudshoorn 2017, which reported that the total cost was similar in women administered 150 IU compared to those given 100 IU (EUR 4714 versus EUR 4622; MD EUR −92, 95% CI −479 to 325).

## ORT-algorithm studies

#### 4. ORT-based algorithm compared to standard dose OR non-ORT algorithm

Five studies were included in this comparison: four comparing an ORT-based algorithm to a standard dose of 150 IU (Nyboe Andersen 2017; Olivennes 2015; Oudshoorn 2017; Popovic-Todorovic 2003), and one comparing an ORT-based algorithm to an algorithm that did not use any ORT (Allegra 2017; Summary of findings for the main comparison). We created one of these trials by merging results from two other trials (three dose comparisons) reported in comparisons one and three (Oudshoorn 2017; Van Tilborg 2017), and we reference it here as Oudshoorn 2017.

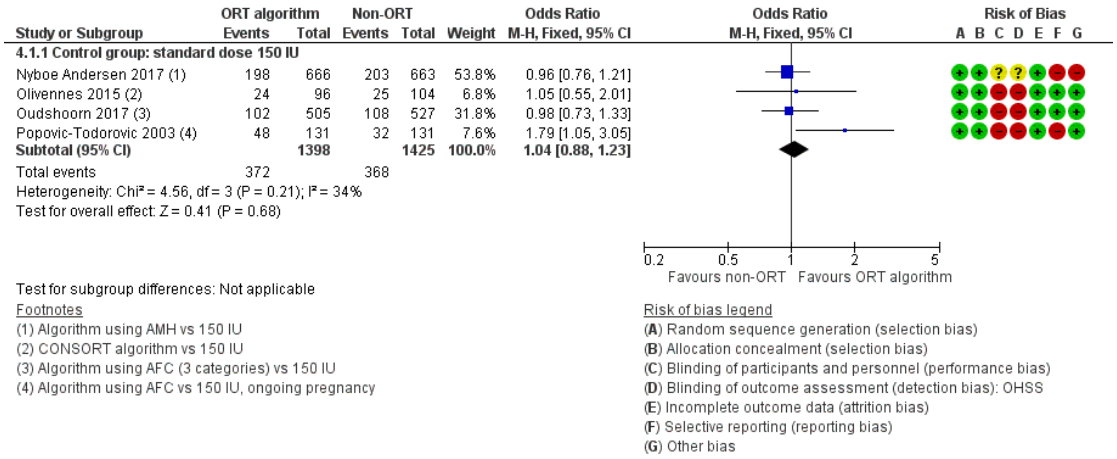
As the trials in this comparison display considerable protocol heterogeneity, with different algorithms tested in each, we would advise the reader to interpret pooled estimates loosely as summaries of the existing studies, rather than as unified estimates of underlying treatment effects.

### Primary outcomes

### *4.1 Live birth or ongoing pregnancy*

Data on live births were available for three studies (Nyboe Andersen 2017; Olivennes 2015; Oudshoorn 2017), and one study reported ongoing pregnancy (Popovic-Todorovic 2003). There was no clear evidence of a difference between the groups in live birth rate (OR 1.04, 95% CI 0.88 to 1.23; N = 2823; 4 studies; $I^2$ = 34%; Analysis 4.1; Figure 9). This suggests that if the chance of live birth with a standard dose is 26%, the chance with dosing based on an ORT-algorithm would be between 24% and 30%. We graded the body of evidence as moderate quality.

**Figure 9. Forest plot of comparison: 4 ORT-based algorithm vs standard dose or non-ORT based algorithm, outcome: 4.1 Live birth or ongoing pregnancy per woman randomised.**




### *4.2 Severe OHSS*

The occurrence of severe OHSS was reported or confirmed following author correspondence for four studies (Allegra 2017;

Olivennes 2015; Oudshoorn 2017; Popovic-Todorovic 2003). A fifth study only provided OHSS for the combined outcome of moderate/severe, and it was not possible to obtain the number of severe events (Nyboe Andersen 2017). We are unable to comment on the effects on severe OHSS, since there was a total of only nine events across two studies, and we graded the body of evidence as low quality; therefore, we refrain from interpreting these results further (Analysis 4.2).

- **ORT-based algorithm versus 150 IU** (Peto OR 0.54, 95% CI 0.14 to 1.99; N = 1494; 3 studies; $I^2$ = 0%).
- **ORT-based algorithm versus non ORT-based algorithm**: no events (Allegra 2017).

**Secondary outcomes**

### *4.3 Clinical pregnancy*

There did not appear to be any substantial difference in the rate of clinical pregnancy when comparing an ORT-based algorithm to a standard dose of 150 IU (OR 0.96, 95% CI 0.82 to 1.13; N = 2823; 4 studies; $I^2$ = 11%), but there was greater uncertainty when comparing to a non-ORT based algorithm (OR 0.88, 95% CI 0.48 to 1.61; N = 194; Analysis 4.3). This suggests that if the chance of clinical pregnancy with a standard dose of 150 IU is 32%, the chance with dosing based on an ORT-algorithm would be between 28% and 35%, and if the chance with a non-ORT based algorithm is 33%, then the chance with an ORT-based algorithm would be between 19% and 45%. We graded the body of evidence as moderate quality.

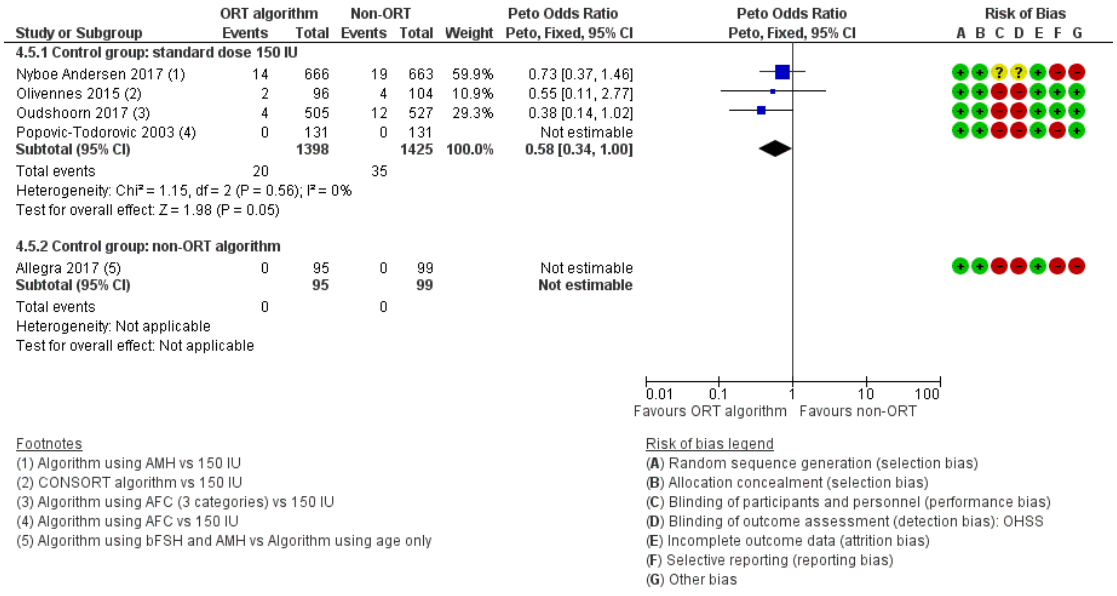
### *4.4 Time to clinical pregnancy*

None of the studies reported this outcome (Oudshoorn 2017 reported the time to ongoing pregnancy, but not time to clinical pregnancy).

### *4.5 Moderate or severe OHSS*

The pooled evidence suggested that the use of an ORT-based algorithm reduced the incidence of moderate or severe OHSS when compared to a standard dose of 150 IU (Peto OR 0.58, 95% CI 0.34 to 1.00; N = 2823; 4 studies; $I^2$ = 0%), but there were no events observed when comparing to a non-ORT algorithm (not estimable, 0 events in each arm, 1 study; Analysis 4.5; Figure 10). This suggests that if the chance of moderate or severe OHSS with a standard dose is assumed to be 2.5%, the chance with dosing based on an ORT-algorithm would be between 0.8% and 2.5%. We graded the body of evidence as low quality.

**Figure 10. Forest plot of comparison: 4 ORT-based algorithm vs standard dose or non-ORT based algorithm, outcome: 4.5 Moderate or severe OHSS.**



### *4.6 Multiple pregnancy in randomized women*

There was insufficient evidence to determine whether there was a difference in the rate of multiple pregnancies when comparing a standard dose of 150 IU to an ORT-based algorithm, although the confidence interval was consistent with small effects in either direction (Peto OR 0.77, 95% CI 0.43 to 1.36; N = 2823; 4 studies; $I^2$ = 0%; Analysis 4.6). This suggests that if the chance of multiple pregnancy with a standard dose is 2.0%, the chance with dosing based on an ORT-algorithm would be between 0.9% and 2.7%. We also calculated the data as multiple pregnancy in women with clinical pregnancy, with similar results (Analysis 4.20). The trial comparing an ORT-based algorithm to a non-ORT based algorithm did not report this outcome (Allegra 2017).

### *4.7 Number of oocytes retrieved per woman randomized*

These are mean differences on the log scale and should not be misinterpreted as numbers of eggs.

All five trials reported this outcome (Analysis 4.7). There is evidence of a higher number of oocytes retrieved in women using an ORT-based algorithm compared to a standard dose of 150 IU (log(MD) oocytes 0.16, 95% CI 0.11 to 0.20; N = 1770; 4 studies, $I^2$ = 99%). However, the heterogeneity is very high, with different studies reporting effects in different directions, indicating that there is scope for substantial variation according to the particular algorithm used and/or the characteristics of the treated population.

The estimate of the effect of ORT-based versus non-ORT-based algorithm suggests that there are probably higher numbers of oocytes for the former compared to the latter (log(MD) oocytes 0.12, 95% CI −0.02 to 0.26, N = 194), although the magnitude of the difference is unclear given the width of the CI.

### *4.8 Poor response to stimulation*

This outcome was available for three studies, which defined a poor response as the retrieval of fewer than five oocytes (Popovic-Todorovic 2003), fewer than nine oocytes (Allegra 2017), and either cycle cancellation for poor response or the retrieval of fewer than five oocytes (Oudshoorn 2017). The pooled effect would be consistent with a small decrease or anything up to a substantial increase in the probability of poor response under an ORT-algorithm compared to standard dosing; however, the heterogeneity is very high (OR 1.16, 95% CI 0.90 to 1.50; N = 1294; 2 studies; $I^2$ = 89%; Analysis 4.8). By contrast, the probability of poor response was reduced under an ORT-algorithm compared to a non-ORT algorithm; however, the magnitude of the reduction is imprecise (OR 0.50, 95% CI 0.27 to 0.92; N = 194; Analysis 4.8; Allegra 2017).

This suggests that if the chance of a poor response with a standard dose is 26%, then the chance with ORT-algorithm dosing would be between 24% and 34%. Further, if the chance of retrieving eight or fewer oocytes with a non-ORT algorithm was 40%, then the chance with ORT-algorithm dosing would be between 16% and 38%.

Another study reported the total number of women experiencing either a poor or hyper-response as an aggregate outcome and declined to provide the number of women experiencing a poor response and hyper-response separately (Nyboe Andersen 2017).

### *4.9 Normal response to stimulation*

Four studies reported this outcome and defined a normal response as retrieval of 5 to 14 oocytes (Popovic-Todorovic 2003), 5 to 15 oocytes (or cycle cancellation for any reason other than a poor or hyper-response) (Oudshoorn 2017), and 8 to 14 oocytes (Allegra 2017; Nyboe Andersen 2017).

The pooled effect suggested an increase in the probability of a normal response to stimulation when using an ORT-algorithm compared to standard dosing (OR 1.22, 95% CI 1.04 to 1.43; N = 2623; 3 studies; $I^2$ = 0%; Analysis 4.9) or non-ORT algorithm dosing (OR 2.13, 95% CI 1.20 to 3.78; N = 194; Analysis 4.9; Allegra 2017). This suggests that if the chance of having a normal response with a standard dose is 45%, then the chance with ORT-algorithm dosing would be between 46% and 54%, and further, if the chance with a non-ORT algorithm was 42%, then the chance with ORT-algorithm dosing would be between 47% and 74%.

### *4.10 Hyper-response to stimulation*

This outcome was available for three studies. All defined a hyper-response as 15 or more oocytes retrieved (Allegra 2017; Popovic-Todorovic 2003; Oudshoorn 2017; Analysis 4.10).

The use of an ORT-algorithm appeared to reduce the number of women with a hyper-response compared to a standard dose of 150 IU; however, heterogeneity was high (OR 0.56, 95% CI 0.42 to 0.76; N = 1294; 2 studies; $I^2$ = 81%). There was insufficient evidence to determine whether there was any difference between the groups when comparing an ORT-based algorithm to a non-ORT based algorithm (OR 0.57, 95% CI 0.25 to 1.31; N = 194; Allegra 2017). This suggests that if the chance of having a hyper-response with a standard dose is 21%, then the chance with ORT-algorithm dosing would be between 9.8% and 16%, and further, if the chance with a non-ORT algorithm was 17%, then the chance with ORT-algorithm dosing would be between 4.9% and 21%.

As described above, we were unable to clarify this outcome, which appears to have been recorded but not reported for one study (Nyboe Andersen 2017).

### *4.11 Cycle cancellations for poor response*

Cycle cancellation due to poor response was an uncommon event, so the effect estimates are imprecise when comparing an ORT-based algorithm to a standard dose of 150 IU (OR 1.19, 95% CI 0.89 to 1.60; N = 2823; 4 studies; $I^2$ = 0%; Analysis 4.11), or a

non-ORT based algorithm (OR 0.14, 95% CI 0.01 to 2.83; N = 194; Analysis 4.11; Allegra 2017). This suggests that if the chance of having a cycle cancellation for poor response with a standard dose is 6.5%, then the chance with ORT-algorithm dosing would be between 5.8% and 9.9%.

### *4.12 Cycle cancellations for hyper-response*

An ORT-based algorithm resulted in fewer cancellations for hyper-response than a standard dose of 150 IU (OR 0.37, 95% CI 0.24 to 0.57; N = 2823; 4 studies; $I^2$ = 3%; Analysis 4.12). This suggests that if the chance of having a cycle cancellation for hyper-response with a standard dose is 5.3%, then the chance with ORT-algorithm dosing would be between 1.3% and 3.1%. However, there was insufficient evidence to determine whether there was any difference in cancellation rates for hyper-response in the study comparing an ORT-based algorithm with a non-ORT based algorithm (OR 0.95, 95% CI 0.40 to 2.27; N = 194; Analysis 4.12; Allegra 2017). This suggests that if the chance of having cycle cancellation for hyper-response with a non-ORT algorithm is 12%, then the chance with ORT-algorithm dosing would be between 5.2% and 24%.

### *4.13 Cycle cancellations for poor or hyper-response*

Cycle cancellation or freeze-all due to hyper-response occurred less often than cancellation for poor response, so the occurrence of cancellation for a poor response contributed more to this analysis. Overall, the pooled evidence suggested a reduced probability of cycle cancellations for either poor or hyper-response using an ORT-based algorithm compared to a standard dose of 150 IU (OR 0.78, 95% CI 0.61 to 1.00; N = 2823; 4 studies; $I^2$ = 0%); however, there was no effect when comparing an ORT-based algorithm with a non-ORT algorithm (OR 0.73, 95% CI 0.32 to 1.69; N = 194; Allegra 2017; Analysis 4.13). This suggests that if the chance of having a cycle cancellation for poor or hyper-response with a standard dose is 12%, then the chance with ORT-algorithm dosing would be between 7.5% and 12%, and further, if the chance with a non-ORT algorithm was 15%, then the chance with ORT-algorithm dosing would be between 5.4% and 23%.

### *4.14 Women with at least one transferable embryo*

Four studies reported this outcome, which refers to the number of women with at least one embryo available to transfer, i.e. either undergoing a fresh transfer, or having a freeze-all (Allegra 2017; Olivennes 2015; Oudshoorn 2017; Popovic-Todorovic 2003; Analysis 4.14). The estimate of the difference in the number of women with embryos available to transfer was imprecise and consistent with nontrivial effects in either direction when comparing an ORT-based algorithm with a standard dose of 150 IU (OR 0.90, 95% CI 0.74 to 1.10; N = 2823; 4 studies; $I^2$ = 0%), or a non-ORT algorithm (OR 1.15, 95% CI 0.57 to 2.33; N = 194; Allegra 2017). This suggests that if the chance of having at least one transferable embryo with a standard dose is 84%, then the chance with ORT-algorithm dosing would be between 80% and 85%, and further, if the chance with a non-ORT algorithm was 79%, then the chance with ORT-algorithm dosing would be between 68% and 90%.

### *4.15 Total dose of FSH*

When comparing an ORT-based algorithm to standard dose of 150 IU, the pooled mean total dose of FSH administered was lower in the ORT-based algorithm group (MD IU −157.00, 95% CI −215.54 to −98.45; N = 1494; 3 studies; $I^2$ = 96%; Analysis 4.15). However, the three studies included in this estimate report very different results for this outcome, and the large effect reported here is largely due to a single trial that reports a much lower total dose in the ORT-algorithm arm (Olivennes 2015). All three trials comparing ORT-based algorithm to 150 IU reported this outcome; however, the doses used by one study were in a drug-specific unit (µg) that we could not pool with the other studies (which used a unit of IU). This study reported a significantly higher average total dose in women in the standard dosing arm (mean (SD) 90.0 µg (25.3) versus 103.7 µg (33.6) P < 0.001; Nyboe Andersen 2017).

In the study comparing an ORT-based algorithm to a non-ORT algorithm, there was no clear evidence of a difference in total dose between the two arms (MD IU−11.00, 95% CI −210.30 to 188.30; N = 194; Analysis 4.15; Allegra 2017).

Readers should treat these results with caution due to handling of censored cycle cancellations.

### *4.16 Duration of FSH administration*

When comparing an ORT-based algorithm to standard dose of 150 IU, the average duration of stimulation in days was slightly longer in the ORT-based algorithm group (MD days 0.23, 95% CI 0.09 to 0.37; N = 2823; 4 studies; $I^2$ = 14%; Analysis 4.16); however, there was no difference in days of FSH exposure when comparing an ORT-algorithm to a non-ORT algorithm (MD days −0.40, 95% CI −0.84 to 0.04; N = 194; Analysis 4.16; Allegra 2017).

These results should be treated with caution due to handling of censored cycle cancellations.

## 5. ORT-based algorithm versus different ORT-based algorithm

Three trials compared different ORT algorithms against each other: AMH versus AFC (Lan 2013), AFC plus AMH versus AFC alone (Magnusson 2017), and AMH plus AFC plus bFSH versus bFSH alone (Tasker 2010).

### 5.1 AMH-based algorithm versus AFC-based algorithm

One trial (N = 348) made this comparison (Lan 2013).

**Primary outcomes**

#### *5.1.1 Live birth or ongoing pregnancy*

The trial did not report this outcome.

#### *5.1.2 Severe OHSS*

There were no events in either study arm (Analysis 5.2).

**Secondary outcomes**

#### *5.1.3 Clinical pregnancy*

The confidence intervals were wide and encompassed the possibility of an effect in either direction (OR 0.82, 95% CI 0.53 to 1.27; N = 348, low-quality evidence; Analysis 5.3). This suggests that if the chance of clinical pregnancy with an AFC algorithm is 39%, then the chance with an AMH algorithm would be between 25% and 45%.

#### *5.1.4 Time to clinical pregnancy*

The trial did not report this outcome.

#### *5.1.5 Moderate or severe OHSS*

Findings were inconclusive, but there appeared to be a higher probability of moderate or severe OHSS in those with an AMH-based dose compared to an AFC-based dose (Peto OR 4.28, 95% CI 0.96 to 19.07; N = 348, very low-quality evidence; Analysis 5.5). This suggests that if the chance of moderate or severe OHSS with dosing based on AFC is 0.6%, then the chance with dosing using AMH would be between 0.6% and 9.9%.

#### *5.1.6 Multiple pregnancy in randomized women*

The confidence interval was consistent with substantial effects in multiple pregnancy rates in either direction (Peto OR 1.21, 95% CI 0.66 to 2.23; N = 348; Analysis 5.6). This suggests that if the chance of multiple pregnancy with dosing based on AFC is 13%, then the chance with dosing using AMH would be between 8.7% and 24%. We also calculated the multiple pregnancy rates per clinical pregnancy, with similar results (Analysis 5.19).

#### *5.1.7 Number of oocytes retrieved per woman randomized*

These are mean differences on the log scale and should not be misinterpreted as numbers of eggs.

There were more oocytes retrieved in women with dose-selection using an AFC algorithm compared to an algorithm using AMH (log(MD) oocytes −0.25, 95% CI −0.37 to −0.13; N = 348; Analysis 5.7).

#### *5.1.8 Poor response to stimulation*

Findings were suggestive of a higher rate of poor response in the arm using the AMH algorithm than in the arm using an AFC algorithm, although the confidence interval crossed the line of no effect (OR 2.25, 95% CI 0.94 to 5.35; N = 348; Analysis 5.8).

#### *5.1.9 Normal response to stimulation*

This outcome was defined as the retrieval of 8 to 12 oocytes. There was no clear difference between the groups in the number of participants experiencing a normal response, but confidence intervals were wide (OR 1.38, 95% CI 0.87 to 2.17; N = 348; Analysis 5.9).

#### *5.1.10 Hyper-response to stimulation*

This outcome was defined as more than 12 oocytes. There were fewer women with more than 12 eggs collected among those allocated to an AMH-based algorithm, compared to an AFC-based algorithm (OR 0.45, 95% CI 0.23 to 0.88; N = 348; Analysis 5.10).

#### *5.1.11 Cycle cancellations for poor response*

We could not draw any conclusions, as the rate of cycle cancellation for poor response was low and therefore the confidence intervals for the effect estimates are imprecise (OR 1.51, 95% CI 0.25 to 9.14; N = 348; Analysis 5.11).

#### *5.1.12 Cycle cancellations for hyper-response*

We could not draw any conclusions, as the rate of cycle cancellation for hyper-response was low, so the confidence intervals for the effect estimates are imprecise (OR 0.54, 95% CI 0.23 to 1.25; N = 348; Analysis 5.12). This suggests that if the chance of cycle cancellation for hyper-response with an AFC algorithm is 9.2%, then the chance with an AMH algorithm would be between 2.3% and 11%.

#### *5.1.13 Cycle cancellations for poor or hyper-response*

We could not draw any conclusions, as the rate of cycle cancellation for poor or hyper-response was low, so the confidence intervals for the effect estimates are imprecise (OR 0.64, 95% CI 0.30 to 1.38; N = 348; Analysis 5.13). This suggests that if the chance of cycle cancellation for either poor or hyper-response with an AFC algorithm is 10%, then the chance with an AMH algorithm would be between 3.3% and 14%.

#### *5.1.14 Women with at least one transferable embryo*

This outcome refers to the number of women who had at least one embryo available for transfer, either for a fresh embryo transfer or for a freeze-all strategy. We could not draw any conclusions, as the confidence intervals are imprecise and consistent with large effects in either direction (OR 1.21, 95% CI 0.36 to 4.03; N = 348; Analysis 5.14). This suggests that if the chance of having at least one transferable embryo with an AFC algorithm is 97%, then the chance with an AMH algorithm would be between 91% and 99%.

#### *5.1.15 Total dose of FSH*

We could not draw any conclusions, as the confidence intervals are imprecise and could be consistent with small effects in either direction (MD IU −178.00, 95% CI −413.88 to 57.88, N = 348; Analysis 5.15). We would urge caution when interpreting this result, as it was not possible to account for censored data arising from cancelled cycles.

#### *5.1.16 Duration of FSH administration*

There was no clear evidence of a difference between the two dosing algorithms in the duration of FSH administration (MD 0.20 days, 95% CI −0.11 to 0.51, N = 348; Analysis 5.16). We would urge caution when interpreting this result as it was not possible to account for censored data arising from cancelled cycles.

### 5.2 AMH plus AFC-based algorithm versus AFC-based algorithm

One study (N = 308) made this comparison (Magnusson 2017).

#### Primary outcomes

#### *5.2.1 Live birth or ongoing pregnancy*

The confidence intervals were wide and encompassed the possibility of an effect in either direction (OR 1.18, 95% CI 0.72 to 1.93; N = 308, moderate-quality evidence; Analysis 6.1; Figure 11). This suggests that if the chance of live birth with a ORT dosing using AFC only is 27%, then the chance with dosing based on AFC and AMH would be between 21% and 42%.

**Figure 11. Forest plot of comparison: 6 AMH + AFC-based algorithm vs AFC-based algorithm, outcome: 6.1 Live birth or ongoing pregnancy.**


#### *5.2.2 Severe OHSS*

Findings were inconclusive, as there were only five events in total (Peto OR 0.68 95% CI 0.12 to 4.00; N = 308, low-quality evidence; Analysis 6.2).

#### Secondary outcomes

#### *5.2.3 Clinical pregnancy*

The confidence intervals were wide and encompassed the possibility of an effect in either direction (OR 1.37, 95% CI 0.84 to 2.23; N = 308, moderate-quality evidence; Analysis 6.3). This suggests that if the chance of clinical pregnancy with ORT dosing using AFC only is 27%, then the chance with dosing based on AFC and AMH would be between 24% and 45%.

#### *5.2.4 Time to clinical pregnancy*

This outcome was not reported.

#### *5.2.5 Moderate or severe OHSS*

The confidence intervals were wide and encompassed the possibility of an effect in either direction (Peto OR 0.85 95% CI 0.26 to 2.83; N = 308, low-quality evidence; Analysis 6.5). This suggests that if the chance of moderate or severe OHSS with a ORT dosing using AFC only is 3.8%, then the chance with dosing based on AFC and AMH would be between 1.0% and 10.0%.

#### *5.2.6 Multiple pregnancy in randomized women*

There were no multiple pregnancies in either study arm (Analysis 6.6; Analysis 6.19).

#### *5.2.7 Number of oocytes retrieved per woman randomized*

These are mean differences on the log scale and should not be misinterpreted as numbers of eggs.

There were fewer oocytes retrieved in women with dose-selection using an AFC plus AMH algorithm compared to an algorithm using AFC only (log(MD) oocytes −0.19, 95% CI −0.31 to −0.07; N = 308; Analysis 6.7).

#### *5.2.8 Poor response to stimulation*

This outcome was defined as the retrieval of fewer than five oocytes. There were substantially more women with a poor response among those in the algorithm using AMH and AFC, compared to that using an algorithm using AFC only (OR 2.82, 95% CI 1.52 to 5.25; N = 308; Analysis 6.8).

#### *5.2.9 Normal response to stimulation*

This outcome was defined as the retrieval of 5 to 12 oocytes. There was no clear evidence of a difference between the groups in the number of participants experiencing a normal response (OR 0.71, 95% CI 0.45 to 1.12; N = 308; Analysis 6.9).

#### *5.2.10 Hyper-response to stimulation*

This outcome was defined as more than 12 oocytes. Findings were consistent with effects in either direction (OR 0.72, 95% CI 0.43 to 1.23; N = 308; Analysis 6.10).

#### *5.2.11 Cycle cancellations for poor response*

We could not draw any conclusions, as the rate of cycle cancellation for poor response was low, so the confidence intervals for the effect estimates are imprecise (OR 1.83, 95% CI 0.53 to 6.40; N = 308; Analysis 6.11). This suggests that if the chance of cycle cancellation for poor response with a ORT dosing using AFC only is 2.6%, then the chance with dosing based on AFC and AMH would be between 1.4% and 14%.

#### *5.2.12 Cycle cancellations for hyper-response*

We could not draw any conclusions, as the rate of cycle cancellation for hyper-response was low, so the confidence intervals for the effect estimates are imprecise (OR 0.50, 95% CI 0.12 to 2.05; N = 308; Analysis 6.12). This suggests that if the chance of cycle cancellation for hyper-response with a ORT dosing using AFC only is 3.8%, then the chance with dosing based on AFC and AMH would be between 0.5% and 7.6%.

#### *5.2.13 Cycle cancellations for poor or hyper-response*

We could not draw any conclusions, as the rate of cycle cancellation for poor or hyper-response was low, so the confidence intervals for the effect estimates are imprecise (OR 1.03, 95% CI 0.42 to 2.55; N = 308; Analysis 6.13). This suggests that if the chance of cycle cancellation for poor or hyper-response with a ORT dosing using AFC only is 6.4%, then the chance with dosing based on AFC and AMH would be between 2.8% and 15%.

#### *5.2.14 Women with at least one transferable embryo*

This outcome refers to the number of women who had at least one embryo available for transfer, either for a fresh embryo transfer or for a freeze-all strategy. We could not draw any conclusions, as the confidence intervals are wide and imprecise and consistent with large effects in either direction (OR 1.04, 95% CI 0.50 to 2.19; N = 308; Analysis 6.14). This suggests that if the chance of having at least one transferable embryo with ORT dosing using AFC only is 90%, then the chance with dosing based on AFC and AMH could be between 81% and 95%.

#### *5.2.15 Total dose of FSH*

We could not draw any conclusions, as the confidence intervals are wide and imprecise and consistent with small effects in either direction (MD IU 81.00, 95% CI −111.93 to 273.93, N = 308; Analysis 6.15).

We would urge caution when interpreting this result, as it was not possible to account for censored data arising from cancelled cycles.

#### 5.2.16 Duration of FSH administration

Participants in the AMH plus AFC-dosing algorithm had a longer duration of stimulation than those in the group using AFC-dosing algorithm (MD days 0.50 days, 95% CI 0.10 to 0.90, N = 308; Analysis 6.16).

We would urge caution when interpreting this result, as it was not possible to account for censored data arising from cancelled cycles.

### 5.3 AMH plus AFC plus bFSH-based algorithm versus bFSH-based algorithm

One study (N = 286) made this comparison (Tasker 2010).

**Primary outcomes**

#### 5.3.1 Live birth or ongoing pregnancy

Findings were inconclusive but were suggestive of a lower event rate in the AMH plus AFC plus bFSH group (OR 0.54, 95% CI 0.28 to 1.04; N = 215, very low-quality evidence; Analysis 7.1). This suggests that if the chance of live birth with dosing based on bFSH is 28%, then the chance with dosing using AMH, AFC and bFSH would be between 10% and 29%. However, readers should treat this result with caution, as the review team extracted the data from individual participant data, which had substantial amounts of missingness.

#### 5.3.2 Severe OHSS

The trial did not report this outcome.

**Secondary outcomes**

#### 5.3.3 Clinical pregnancy

The clinical pregnancy rate appeared to be lower in the participants having dose-selection based on an algorithm using AMH, AFC and bFSH (OR 0.51, 95% CI 0.28 to 0.93; N = 215, very low-quality evidence; Analysis 7.3). This suggests that if the chance of clinical pregnancy with dosing based on bFSH alone is 36%, then the chance with dosing using AMH, AFC, and bFSH would be between 14% and 35%.

#### 5.3.4 Time to clinical pregnancy

The trial did not report this outcome.

#### 5.3.5 Moderate or severe OHSS

The trial did not report this outcome.

#### 5.3.6 Multiple pregnancy in randomized women

The trial did not report this outcome.

#### 5.3.7 Number of oocytes retrieved per woman randomized

There was no clear difference between the groups in the average number of oocytes retrieved (log(MD) oocytes −0.20, 95% CI −0.81 to 0.41; N = 215; Analysis 7.7).

#### 5.3.8 Poor response to stimulation

This outcome was defined as the retrieval of fewer than five oocytes. We could not draw any conclusions, as the confidence intervals are wide and imprecise (OR 1.46, 95% CI 0.77 to 2.79; N = 215; Analysis 7.8).

#### 5.3.9 Normal response to stimulation

This outcome was defined as the retrieval of 5 to 14 oocytes. There was no clear evidence of a difference between the groups (OR 0.75, 95% CI 0.42 to 1.35; N = 215; Analysis 7.9).

#### 5.3.10 Hyper-response to stimulation

This outcome was defined as retrieval of more than 14 oocytes. We could not draw any conclusions, as the confidence intervals are wide and imprecise (OR 0.93, 95% CI 0.45 to 1.93; N = 215; Analysis 7.10).

#### 5.3.11 Cycle cancellations for poor response

The trial did not report this outcome.

#### 5.3.12 Cycle cancellations for hyper-response

The trial did not report this outcome.

#### 5.3.13 Cycle cancellations for poor or hyper-response

The trial did not report this outcome.

#### *5.3.14 Women with at least one transferable embryo*

The trial did not report this outcome.

#### *5.3.15 Total dose of FSH*

We could not draw any conclusions, as the confidence intervals are wide and consistent with small effects in either direction (MD IU −148.00, 95% CI −433.61 to 137.61, N = 215; Analysis 7.15). We would urge caution when interpreting this result, as it was not possible to account for censored data arising from cancelled cycles.

#### *5.3.16 Duration of FSH administration*

There was no clear evidence of a difference between the groups in the duration of FSH administration (MD 0.00 days, 95% CI −0.60 to 0.60, N = 215; Analysis 7.16). We would urge caution when interpreting this result, as it was not possible to account for censored data arising from cancelled cycles.

### Sensitivity analyses

We stated in the Methods that we would conduct sensitivity or subgroup analyses by excluding studies with various characteristics. As the evidence was limited for most outcomes and associated with significant imprecision, most of the planned sensitivity analyses are moot; if we start with an imprecise estimate, and then take some data away, nothing additional can be learned. If we switch to a random-effects analysis, the estimates and confidence limits for live birth change by no more than 0.01. In addition, we have presented pooled estimates for our ORT-algorithm studies. However, these are intended not to represent estimates of underlying treatment effects, but rather summaries of the results of the individual studies. The considerable protocol variation between ORT-algorithm studies precludes effect estimation. Contrary to popular belief, this in turn precludes random-effects meta-analysis.

## ADDITIONAL SUMMARY OF FINDINGS [Explanation]

**Anticipated low responders: higher compared to lower dose of FSH for women undergoing IVF/ICSI**

**Patient or population**: women undergoing IVF/ICSI who are anticipated to have a low-response to stimulation based on one or more ORT measure
**Setting**: hospital or fertility clinic
**Intervention**: higher dose of FSH
**Comparison**: lower dose of FSH

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Evidence summary |
|---|---|---|---|---|---|---|
| | Risk with lower dose | Risk with higher dose | | | | |
| **Live birth or ongoing pregnancy** | | | | | | |
| 300/450 IU vs 150 IU | 109 per 1000 | 80 per 1000 (38 to 162) | OR 0.71 (0.32 to 1.58) | 286 (2 RCTs) | ⊕⊕◯◯ Low[a,b] | It was difficult to determine whether the different doses impacted on rates of live birth or ongoing pregnancy, therefore there is no evidence to suggest increased dosing in low responders is beneficial |
| 400/450 IU vs 300 IU | 161 per 1000 | 129 per 1000 (35 to 380) | OR 0.77 (0.19 to 3.19) | 62 (1 RCT) | | |
| 600 IU vs 450 IU | 108 per 1000 | 139 per 1000 (79 to 234) | OR 1.33 (0.71 to 2.52) | 356 (1 RCT) | | |
| **Ovarian hyperstimulation syndrome (OHSS)** | | | | | | |
| 300/450 IU vs 150 IU | *Severe* 0 per 1000 | 0 per 1000 (0 to 0) | Not estimable | 286 (2 RCTs) | ⊕◯◯◯ Very low[a,c] | It was not possible to determine whether the different doses impacted on rates of OHSS, as the event rates were too low |
| | *Moderate or severe* 0 per 1000 | 0 per 1000 (0 to 0) | Not estimable | 286 (2 RCTs) | | |
| 400/450 IU vs 300 IU | *Severe* 0 per 1000 | 0 per 1000 (0 to 0) | Not estimable | 62 (1 RCT) | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | *Moderate or severe*<br>0 per 1000 | 0 per 1000<br>(0 to 0) | Not estimable | 62<br>(1 RCT) | | |
| 600 IU vs 450 IU | *Severe*<br>0 per 1000 | 0 per 1000<br>(0 to 0) | Not estimable | 356<br>(1 RCT) | | |
| | *Moderate or severe*<br>0 per 1000 | 0 per 1000<br>(0 to 0) | OR 7.23<br>(0.14 to 364) | 356<br>(1 study) | | |
| **Clinical pregnancy** | | | | | | |
| 300/450 IU vs 150 IU | 184 per 1000 | 101 per 1000<br>(53 to 184) | OR 0.50<br>(0.25 to 1.00) | 286<br>(2 RCTs) | ⊕⊕◯◯<br>Low[a,b] | It was difficult to determine whether the dose differences impacted on rates of clinical pregnancy, therefore there is no evidence to suggest increased dosing in low responders is beneficial, and it may even be detrimental compared to a dose of 150 IU |
| 400/450 IU vs 300 IU | 127 per 1000 | 109 per 1000<br>(37 to 282) | OR 0.84<br>(0.26 to 2.69) | 110<br>(2 RCTs) | | |
| 600 IU vs 450 IU | 159 per 1000 | 177 per 1000<br>(111 to 274) | OR 1.14<br>(0.66 to 1.99) | 356<br>(1 RCT) | | |

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).
__CI__: confidence interval; __FSH__: follicle-stimulating hormone; __ICSI__: intracytoplasmic sperm injection; __IU__: international units; __IVF__: in vitro fertilisation; __OHSS__: ovarian hyperstimulation syndrome; __OR__: odds ratio; __ORT__: ovarian reserve test; __RCT__: randomised controlled trial.

**GRADE Working Group grades of evidence**
**High certainty**: we are very confident that the true effect lies close to that of the estimate of the effect.
**Moderate certainty**: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
**Low certainty**: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
**Very low certainty**: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect

[a]Downgraded one level for serious risk of bias associated mainly with performance bias due to lack of blinding and/or selective reporting.

[b]Downgraded one level for serious imprecision associated with small number of events.
[c]Downgraded two levels for very serious imprecision associated with very small number of events

**Anticipated normal responders: higher compared to lower dose of FSH for women undergoing IVF/ICSI**

**Patient or population**: women undergoing IVF/ICSI who are anticipated to have a normal response to stimulation based on at least one ORT measure
**Setting**: hospital or fertility clinic
**Intervention**: higher dose of FSH
**Comparison**: lower dose of FSH

| Comparison | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Evidence summary |
|---|---|---|---|---|---|---|
| | Risk with lower dose | Risk with higher dose | | | | |
| **Outcome:live birth or ongoing pregnancy** | | | | | | |
| 200 IU vs 100 IU | 204 per 1000 | 184 per 1000 (127 to 258) | OR 0.88 (0.57 to 1.36) | 522 (2 RCTs) | ⊕⊕◯◯ Low[a,b] | It was difficult to determine whether the dose differences impacted on rates of live birth or ongoing pregnancy, therefore there is no evidence to suggest increased dosing in normal responders is beneficial |
| 225/200 IU vs 150 IU | 193 per 1000 | 198 per 1000 (120 to 308) | OR 1.03 (0.57 to 1.86) | 277 (1 RCT) | | |
| 300 IU vs 225 IU | 397 per 1000 | 300 per 1000 (174 to 465) | OR 0.65 (0.32 to 1.32) | 135 (1 RCT) | | |
| **Outcome:ovarian hyperstimulation syndrome (OHSS)** | | | | | | |
| 200 IU vs 100 IU | *Severe* 4 per 1000 | 1 per 1000 (0 to 26) | OR 0.14 (0.00 to 6.96) | 522 (2 RCTs) | ⊕◯◯◯ Very low[a,c] | It was impossible to determine whether the dose differences impacted on rates of OHSS |
| | *Moderate or severe* 31 per 1000 | 19 per 1000 (7 to 56) | OR 0.62 (0.21 to 1.87) | 522 (2 RCTs) | | |
| 225/200 IU vs 150 IU | *Severe* 8 per 1000 | 8 per 1000 (2 to 39) | OR 1.00 (0.20 to 5.20) | 740 (4 RCTs) | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | *Moderate or severe*<br>27 per 1000 | 32 per 1000<br>(14 to 73) | OR 1.21<br>(0.51 to 2.85) | 740<br>(4 RCTs) | | |
| 300 IU vs 225 IU | *Severe*<br>15 per 1000 | 2 per 1000<br>(0 to 94) | OR 0.14 (0.00 to 6.92) | 135 (1 RCT) | | |
| | *Moderate or severe*<br>44 per 1000 | 30 per 1000<br>(5 to 156) | OR 0.67<br>(0.11 to 3.99) | 135<br>(1 study) | | |
| **Outcome:clinical pregnancy** | | | | | | |
| 200 IU vs 100 IU | 202 per 1000 | 179 per 1000<br>(113 to 274) | OR 0.86 (0.73 to 1.31) | 330 (1 RCT) | ⊕⊕◯◯<br>Low[a,b] | It was difficult to determine whether the dose differences impacted on rates of clinical pregnancy, therefore there is no evidence to suggest increased dosing in normal responders is beneficial |
| 225/200 IU vs 150 IU | 236 per 1000 | 232 per 1000<br>(184 to 288) | OR 0.98<br>(0.75 to 1.33) | 1037<br>(5 RCTs) | | |
| 300 IU vs 225 IU | 441 per 1000 | 418 per 1000<br>(266 to 587) | OR 0.91<br>(0.46 to 1.80) | 135<br>(1 RCT) | | |

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).
__CI__: confidence interval; __FSH__: follicle-stimulating hormone; __ICSI__: intracytoplasmic sperm injection; __IU__: international units; __IVF__: in vitro fertilisation; __OHSS__: ovarian hyperstimulation syndrome; __OR__: odds ratio; __ORT__: ovarian reserve test; __RCT__: randomised controlled trial.

**GRADE Working Group grades of evidence**
**High certainty**: we are very confident that the true effect lies close to that of the estimate of the effect.
**Moderate certainty**: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
**Low certainty**: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
**Very low certainty**: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect

[a]Downgraded one level for serious risk of bias associated mainly with performance bias due to lack of blinding and/or selective reporting and/or selection bias due to unclear methods of randomisation.
[b]Downgraded one level for serious imprecision associated with small number of events.

[c]Downgraded two levels for very serious imprecision associated with very small number of events.

**Anticipated high responders: higher compared to lower dose of FSH for women undergoing IVF/ICSI**

**Patient or population**: women undergoing IVF/ICSI who are anticipated to have a hyper-response to stimulation based on at least one ORT measure
**Setting**: hospital or fertility clinic
**Intervention**: higher dose of FSH
**Comparison**: lower dose of FSH

| Comparison | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Evidence summary |
|---|---|---|---|---|---|---|
| | Risk with lower dose | Risk with higher dose | | | | |
| **Outcome:live birth or ongoing pregnancy** | | | | | | |
| 150 IU vs 100 IU | 255 per 1000 | 251 per 1000 (184 to 333) | OR 0.98 (0.66 to 1.46) | 521 (1 RCT) | ⊕⊕◯◯ Low[a,b] | It was difficult to determine whether the dose differences impacted on rates of live birth or ongoing pregnancy, therefore there is no evidence to suggest increased dosing in hyper responders is beneficial |
| **Outcome:ovarian hyperstimulation syndrome (OHSS)** | | | | | | |
| 150 IU vs 100 IU | *Severe* 16 per 1000 | 11 per 1000 (3 to 48) | OR 0.72 (0.16 to 3.19) | 521 (1 RCT) | ⊕◯◯◯ Very low[a,c] | It was not possible to definitively say whether dose differences impacted on rates of OHSS, but there could be a reduction with lower doses |
| | *Moderate or severe* 16 per 1000 | 36 per 1000 (13 to 96) | OR 2.31 (0.80 to 6.67) | 521 (1 RCT) | | |
| **Outcome:clinical pregnancy** | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 150 IU vs 100 IU | 275 per 1000 | 301 per 1000 (228 to 386) | OR 1.14 (0.78 to 1.66) | 521 (1 RCT) | ⊕⊕◯◯ Low[a,b] | It was difficult to determine whether the dose differences impacted on rates of clinical pregnancy, therefore there is no evidence to suggest increased dosing in hyper responders is beneficial |

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).

**CI**: confidence interval; **FSH**: follicle-stimulating hormone; **ICSI**: intracytoplasmic sperm injection; **IU**: international units; **IVF**: in vitro fertilisation; **OHSS**: ovarian hyperstimulation syndrome; **OR**: odds ratio; **ORT**: ovarian reserve test; **RCT**: randomised controlled trial.

**GRADE Working Group grades of evidence**

**High certainty**: we are very confident that the true effect lies close to that of the estimate of the effect.

**Moderate certainty**: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

**Low certainty**: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.

**Very low certainty**: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect

[a]Downgraded one level for serious risk of bias associated mainly with performance bias due to lack of blinding and/or selective reporting.

[b]Downgraded one level for serious imprecision associated with small number of events.

[c]Downgraded two levels for very serious imprecision associated with very small number of events.

# DISCUSSION

## Summary of main results

### Direct dose comparison studies

#### Predicted low responders

This review included five studies in predicted low responders, each testing a different combination of FSH doses against each other. Four trials reported the primary outcome 'live birth or ongoing pregnancy', and there was insufficient evidence to determine whether there was any difference in the probability of this outcome when women were randomized to multiple doses. For example, if we assume a birth rate of 11% in women treated with 150 IU, then the results of one analysis would be consistent with live birth rates as low as 3.8% or as high as 16.2% in the higher dose (300/450 IU) arm. The effect on cumulative live birth remains unknown, as it was only available for one trial (Van Tilborg 2017).

All five studies reported clinical pregnancy, and the pooled results of one comparison suggest that women administered 150 IU had a higher chance of achieving pregnancy than women administered 300 or 450 IU (albeit with imprecision). This did not translate into an increase in live birth rate, however, since in one of the trials 10 of 24 pregnancies in the lower dosing arm were subsequently lost (Van Tilborg 2017). This may or may not be attributable to chance. In the other two comparisons, sample sizes were too small to draw a conclusion.

There were no events of severe OHSS, the primary safety outcome of this review. Moreover, only one case of moderate OHSS occurred. Therefore we are unable to make any inferences about dose effects on this outcome. However, OHSS is expected to occur only very rarely in a population of women with low predicted ovarian response. Our best available surrogates are cancellation for hyper-response (or freeze-all) and excessive response to stimulation. Although the event rates for the former are too low to be useful, it seems reasonably clear that using 150 IU instead of higher doses reduces hyper-response in this subgroup, as Van Tilborg 2017 reported. However, this must be balanced against the probability of cancellation for poor response, which was higher in the lower dose arms of this trial. The rate of normal response favoured the use of a higher dose of 450 IU.

Only one of the five studies in this cohort had a point estimate indicating a benefit of increasing dose on birth, such that, on balance, it appears unlikely that increasing dose will improve outcomes of live birth or OHSS in this population.

#### Predicted normal responders

There were nine studies included in this comparison, eight of which were available for pooling in meta-analysis. One could not be compared to the rest, as the type of FSH used was expressed in a different unit. In a similar trend to that seen for the low responder group, there was no observed influence from higher or lower doses on the probability of live birth, OHSS or clinical pregnancy. However the estimates were imprecise and remain consistent with small effects in either direction.

It was difficult to assess the five-arm trial for dose-response effects on clinical outcomes, as the numbers of participants and event rates were low.

As with predicted low responders, there is evidence of effects of increasing dose on preclinical outcomes such as oocyte yield and cycle cancellation, but the evidence reported in this review does not suggest that these effects translate into differences in clinical outcomes such as live birth or clinical pregnancy (discussed further below).

#### Predicted high responders

There were only two studies in this comparison, one trial comparing 150 IU versus 100 IU in women with AFC or more than 15, and a five-arm study in women with AMH between 15 pmol/L and 55 pmol/L comparing five different doses of FSH. There was no observed effect of dose reduction below 150 IU on the probability of live birth or clinical pregnancy. However, 100 IU did seem to reduce the occurrence of moderate or severe OHSS compared to 150 IU. This observed benefit from lower doses may be weighed against the disadvantages, as a reduced dose also led to the retrieval of fewer oocytes, an increased chance of cancellation for poor response, more overall cancellations, and fewer women with at least one embryo for transfer; however, these effects did not translate into a reduced probability of clinical outcomes, which could be due to a number of reasons including lack of power, as discussed further below. The event rates in the five-arm trial are consistent with the treatment effects reported in this study.

### ORT-algorithm studies

These studies were divided into two comparisons: those comparing an ORT-based algorithm to a standard dose of 150 IU or another algorithm which did not use ORT, and those comparing two different ORT-based algorithms.

#### ORT-based algorithm versus standard dose or non-ORT based algorithm

There were five trials comparing ORT-based algorithms to a standard dose of 150 IU or an algorithm not using ORT, and these studies were subgrouped according to whether the control arm was a standard dose of 150 IU or a non-ORT based algorithm. Pooling within subgroups was conducted with the caveat that the pooled effects should be interpreted as summaries of the effects estimated in the included studies, rather than an estimate of a distinct underlying treatment effect, as the individual ORT algorithms were different.

When comparing ORT algorithms with 150 IU, on aggregate the studies probably rule out large advantages or disadvantages of ORT-based algorithms on live birth. Point estimates for three of four studies were close to the line of null effect. For each of the two large studies in the comparison, if we were to assume a control group live birth rate of around 20%, estimated treatment group rates would not differ by more than four or five percentage points in either direction (i.e. between 15% and 25%). One trial is somewhat discordant, if not completely inconsistent, with the others, suggesting a benefit of indeterminate magnitude when using an ORT algorithm (Popovic-Todorovic 2003). It is not possible to know whether the apparent difference in this trial proceeds from genuine superiority of the particular algorithm tested compared to the others, from effects of performance bias or selective reporting, or from chance. Further, as this trial was conducted over a decade earlier than the other trials in this comparison, the difference may be a consequence of changes in the general IVF population and in IVF techniques over time. To the extent that we consider the pooled estimate to represent a meaningful summary of this heterogeneous assortment of trials, birth rates under ORT algorithms do not appear to differ meaningfully from those resulting from a uniform fixed 150 IU dose. There were too few cases of severe OHSS to discuss effects on this outcome. If we extended our definition to include both moderate and severe cases, however, then ORT algorithms were consistently associated with reduced OHSS. This subjective outcome might be particularly prone to performance and detection biases arising from a lack of blinding.

Only one study compared an ORT-based algorithm with a non-ORT algorithm, which used only age. The authors found that the ORT algorithm increased the number of oocytes retrieved, reduced the probability of a poor response, and increased the probability of a normal response (with the caveat that the definition of 'normal response' was not prespecified, which probably would have changed the inference of the trial). However, there were insufficient events to determine effects on either pregnancy or OHSS, with no moderate or severe cases in either arm.

#### ORT-based algorithm versus different ORT-based algorithm

Three studies were included in the comparison of different approaches to ORT-based dosing (e.g. AMH versus AFC).

There was insufficient evidence of differences in live birth rate between an AFC-only algorithm and another using both AMH and AFC (Magnusson 2017), although the results would be consistent with advantages of either. Any possible disadvantage of adding AMH to the algorithm would be small (and counterintuitive), however. The clinical pregnancy results supported these observations. There were too few moderate or severe OHSS events to assess which algorithm was safer in this regard.

Another study suggested a disadvantage of adding AMH and AFC to basal FSH to select dose, in relation to clinical pregnancy and birth (Tasker 2010). These results constitute very weak evidence, however, as we pieced them together from an incomplete data set with a higher portion of missing data, and therefore we do not interpret them further.

A third study compared AMH and AFC-based dose selection. Live birth was not reported, and the clinical pregnancy estimate was imprecise (Lan 2013). However, the upper confidence limit suggested at most a small advantage of AMH compared to AFC in this regard, and potentially a large disadvantage. Investigators did not observe any severe OHSS, although there may have been an advantage of indeterminate magnitude of AFC in relation to moderate OHSS; however, this was based on only a few events.

Across all comparisons, we found no impact of dosing on the probability of multiple pregnancy, although estimates were too imprecise to rule anything out.

## Overall completeness and applicability of evidence

### Applicability of the evidence

There were 20 trials included in this review investigating the utility of tailoring FSH doses based on individual ORT measures. Unfortunately, the included studies varied in their design, restricting meta-analysis to subgroups within comparisons, with the caveat that these pooled estimates are unlikely to represent unified underlying treatment effects due to clinically heterogeneous trials. For example, in comparison one, all five included studies compared unique combinations of doses of FSH.

Across the first three comparisons investigated (direct dose comparison studies investigating effects of different doses in different populations), there were apparent effects of increasing dose on the 'upstream' outcomes of IVF, including:

- oocyte yield, in terms of the average number of oocytes retrieved (although effect sizes were variable and often reasonably modest - e.g. an MD of 0.3 on the log scale translates to about a 1.3 factor increase, or the difference between six and eight eggs);
- the number of women categorised as having either a poor (e.g. fewer than 5 oocytes), normal, or hyper-response (e.g. more than 15 oocytes);
- the chance of cycle cancellation for either a poor or hyper-response; and
- the probability of women having at least one embryo available for transfer (or freezing).

Importantly, demonstrable effects on these outcomes fall short of demonstrating effects on clinical outcomes. In the (larger) ORT-algorithm studies, we also saw apparent effects on upstream outcomes but were able to get a slightly better idea of the (lack of) effects on live birth compared to uniform 150 IU dosing. It is impossible to establish how much of the observed variation was attributable to the particular algorithm used in each trial, however.

There are a number of possible reasons as to why effects on these upstream outcomes did not apparently translate into effects on clinical outcomes, namely live birth and pregnancy.

First of all, detection of treatment effects on relatively common events (such as having an embryo available for transfer) or on continuous outcomes (such as average number of oocytes retrieved) is possible with smaller sample sizes than those required to detect treatment effects for rarer binary outcomes, such as live birth. Underpowering for the primary outcomes of this review in the individual direct dose comparison studies is not surprising, as many of these studies appear to have been designed to answer mechanistic research questions, so they based sample size on preclinical outcomes such as the number of oocytes retrieved (e.g. Arce 2014; Harrison 2001; Klinkert 2005; Lefebrve 2015). While ORT-algorithm studies were larger, they were often powered on the basis of reasonably large effect sizes for the outcome of pregnancy or live birth, such as 15% or more (e.g. Magnusson 2017), or else they were designed only to be pilot studies (e.g. Lan 2013).

Secondly, most included studies only assessed the outcome of pregnancy (and birth) following a fresh embryo transfer, and only a few trials reported cumulative rates (following the transfer of a fresh and all frozen embryos). This is important because it could be that an increased dose of FSH leads to higher freeze-all rates for excessive response, as was observed in a number of comparisons among predicted normal and high responders. In most of the included studies, women with freeze-all cycles were not given an opportunity to conceive in the study because the result of the first frozen transfer was not captured. Further, a higher dose of FSH may yield more oocytes and therefore more embryos for frozen transfer (in addition to any fresh transfer). However, the utility of having extra candidate embryos for transfer remains uncertain, since the included studies did not test the probability of these leading to a live birth in a frozen cycle. It was for this reason that this review captured the outcome 'women with at least one transferable embryo' which was a count of the women who either underwent fresh embryo transfer or who had a freeze-all cycle (for any reason). However, the limitations remain that this outcome does not consider the *number* of transferable embryos each women had, and further, that an outcome related to the number of embryos remains a surrogate, rather than a clinically important, outcome. This review did not capture the outcome of 'number of transferable embryos', which may have provided an indication of whether higher observed rates of hyper-response and freeze-all in the higher dosing arms resulted in a higher number of embryos available for transfer. However, ultimately, availability of more embryos (for transfer or freezing) does not definitively lead to a higher chance of pregnancy and live birth (as discussed below).

The above two points suggest that there may be an unobserved but true effect from individual dose selection on the outcome of live birth. It is worth emphasising that there may not be any association between upstream outcomes (e.g. increased number of oocytes or cycle cancellation) and the probability of pregnancy or live birth. Previous observational research has suggested that the optimum number of oocytes retrieved during an IVF cycle is between 5 and 15, and the retrieval of 20 or more oocytes reduces the probability of pregnancy from a fresh IVF cycle (Sunkara 2011). However, it does not necessarily follow that, for example, if a women has four oocytes collected in her first cycle, that she will benefit from an increased number of oocytes retrieved in her second cycle (which may or may not be achieved by increasing her FSH dose). This correlation is only an observation that women who do obtain between 5 and 15 oocytes at retrieval have a higher probability of pregnancy and live birth.

Another possible reason for the observed effect on upstream outcomes but not clinical outcomes is that the increased FSH dose, possibly in combination with excess oestrogen production in the ovaries, can have a detrimental effect on the quality of the endometrium, reducing its receptivity and therefore the probability of implantation (Bourgain 2003; Kolibianakis 2002). Such suggestions have led to the increasing implementation of the freeze-all strategy as routine, whereby all embryos are frozen for transfer in a subsequent non-stimulated cycle (Wong 2017). Indeed, the aforementioned observational study excluded cycles in which a freeze-all strategy was employed (Sunkara 2011).

Moreover, there may be an additional effect of increased FSH dosing on the oocytes themselves, and subsequently on the number and quality of embryos available for transfer (or freezing). In one study included in this review, increasing doses of FSH led to increased oocyte yield but simultaneously reduced the fertilisation rate, resulting in fewer blastocysts per oocyte (Arce 2014). Similar trends were apparent in other included studies (Harrison 2001; Hoomans 2002). Several studies have demonstrated an effect of increasing FSH dose on the oocyte quality and aneuploidy rate in animal models (McGowan 1985; Roberts 2005; Sugano 1997). However, research has not consistently detected the same effect in humans (La Marca 2017; Labarta 2017; Ting 2009). This reinforces the need for caution regarding the use of surrogate outcomes such as the number of oocytes retrieved or the number of women achieving a 'normal response', which is commonly used as a primary outcome in IVF trials (Arce 2014; Bastu 2016; Jayaprakasan 2010; Klinkert 2005; Lan 2013; Lefebrve 2015; Magnusson 2017).

Finally, the role of performance bias is also a relevant consideration, as many of these upstream outcomes (e.g. number of oocytes retrieved, cycle cancellation) can be prone to bias arising from the clinician's knowledge of trial allocation. This could contribute to the treatment effects observed for these outcomes (however, this is equally likely to affect the primary outcome of live birth).

### Applicability of different trial designs

There were two different trial designs included in this review:

1. Direct dose comparison studies: recruiting a subgroup of women based on an ORT measure (e.g. AFC of 0 to 7) and

randomising them to two or more doses of FSH (e.g. 450 IU versus 150 IU).

2. ORT-algorithm studies: recruiting a more general group of women, and randomising them to dose selection based on their individual ORT measure compared to either a standard dose, an algorithm not using an ORT, or a different ORT-based algorithm (e.g. AMH-based algorithm versus 150 IU).

Direct dose comparison studies answer the following question: in women within a given ORT range, do different doses result in better outcomes? The interpretation of these studies is simple and directly applicable to practice. ORT-algorithm studies instead answer the question of whether an algorithm that assigns women to doses of FSH according to their ORT is better than giving everyone the same dose (or alternatively, than using a different algorithm). The general applicability of ORT-algorithm trial results to practice is less clear. Many of the ORT-algorithm studies compared an ORT-based algorithm to the consistent administration of 150 IU. In practice, clinicians often tailor the FSH dose to some extent, based on characteristics such as age, presence of PCOS, and response to stimulation in previous cycles. Only one study, Allegra 2017, considered the value of using ORT to tailor the dose, compared to tailoring the dose without ORT (specifically on the basis of age). The review therefore provides limited information about the comparative effectiveness of ORT dosing compared to routine practice. Moreover, all of the ORT-algorithm studies excluded women with PCOS, who are known to be at higher risk of hyper-response (Aljawoan 2012), and one study specifically excluded women who had previous poor or hyper-response to ovarian stimulation (Olivennes 2015). Many of the study cohorts are therefore unlikely to be representative of typical subfertile populations. However, the exclusion of PCOS women from these trials is sensible, as they are known to respond variably and usually excessively to ovarian stimulation. This matters in the context of this review, since there are reasons to expect individualisation to be more or less effective depending on a woman's ovarian reserve (Arce 2014). On balance, however, we decided not to downgrade the assessed quality of the evidence due to this apparent indirectness, as the extent of the indirectness varies depending on the clinical practice within different fertility clinics.

### Heterogeneity in studied populations

We also note that ORT values of women classified as predicted low, normal, or high responders varied between the studies. For example in the low-responder population the median AFC in two sub-studies was 3 (Bastu 2016; Van Tilborg 2017), while in another two studies it was 9 (Lefebrve 2015). Further, we permitted the definition of a population subgroup (low, normal, or high responders) on the basis of one or more ORT measures, specifically: AMH, AFC, or bFSH. Extensive literature indicates that these measures are not equally useful for predicting response to stimulation; AMH appears be superior to AFC, and both appear to be superior to bFSH (Broekmans 2006; Broer 2013a; Broer 2013b; Broer 2014). Additionally, and as mentioned below, we included any trials comparing different FSH doses, so long as the trial population could be categorised as predicted: low, normal, or high responders based on available ORT data. This may have led to the inclusion of studies with more clinically heterogeneous populations. For example, a study that we included in the comparison in 'predicted low responders' because all bFSH of more than 8.5 IU/L (for example, Harrison 2001) would include more clinically diverse women than a study defining their trial population as women with AFC of 0 to 7 (Van Tilborg 2017). Indeed, most of the trials in the normal responder category defined their trial population on the basis of bFSH alone, so these trials are likely to contain a mixture of normal and high responders as per AFC and AMH (however, we do not have this information). Further, the different ORT measures have other strengths and weaknesses other than their predictive capacities. The strength of AMH is the investigator-independence and stability, and limitations include the use of different assays and difficulty in translating values between assays (Broer 2014). The strength of AFC is that it is easily measured during other routine ultrasound assessment and thus has limited additional cost; however, the weakness is the potential for inter-observer variation.

Often, this clinical and design heterogeneity between trials might be expected to strengthen the external validity of the review. In this case, however, since we usually observed only one trial comparing any particular pair of interventions, differences between populations serve as an unwelcome source of between-study confounding. Despite this, we made a post hoc decision to pool a number of these trials, and this may contribute to the observed statistical heterogeneity in several of the review outcomes, especially average number of oocytes retrieved and total dose and duration of FSH administration. Moreover, this is a research question where the ORT values used to classify women are crucial, since it is the utility of these thresholds for the purpose of matching people to doses that is under investigation. This heterogeneity is a clear barrier to making a unified judgement regarding the effectiveness and safety of ORT dosing.

### Uncertainty regarding safety

There is an expectation that ORT-based dosing might reduce incidence of OHSS. Because event rates were so low, this review provides little information about this outcome, especially for the case of severe OHSS. However, it appears that using ORT algorithms may, in general, reduce the incidence of OHSS. However, the effect size remains unclear, and high-quality analyses of large observational data sets may be needed to fill this gap.

## Quality of the evidence

Evidence quality ranged from very low to moderate. The main limitations were imprecision and risk of bias associated with lack of blinding.

We conducted GRADE assessments for the main review comparisons; for effects of increasing dose in predicted low, normal, and high responders; and for ORT algorithm versus dosing without ORT. We did this separately for the outcomes live birth, moderate and severe OHSS, and clinical pregnancy. Our assessments were generally consistent across the comparisons: we considered quality of evidence to be low for live birth for all comparisons except for comparison 4 (ORT algorithms versus dosing without ORT), where we judged the evidence to be of moderate quality. We downgraded live birth in all comparisons for high risk of bias in the individual studies. We were particularly concerned about performance bias due to a lack of blinding in many of the studies. Patients are monitored during ovarian stimulation, and clinicians make judgements about dose adjustments and about whether or not to cancel IVF cycles that are not going well. There is clearly scope for knowledge of treatment assignment to influence these decisions. There was some evidence of this in one included trial, where a greater proportion of under-performing cycles were cancelled in the 150 IU compared to the individualised (higher dose) arm (Van Tilborg 2017). The trial authors hypothesised that the treating clinicians were more likely to cancel the cycle of women on a lower dose of FSH than in women with a higher FSH dose. In most comparisons we further downgraded for imprecision; confidence intervals were so wide as to be consistent with a plurality of plausible scenarios.

We considered the evidence to be of very low quality for the outcome of OHSS (moderate and severe) for all comparisons, since there were very few cases, so the effects of individualised dosing on this outcome were often unclear. However, using an ORT-based algorithm does appear to reduce the probability of moderate or severe OHSS compared to a standard dose of 150 IU.

We considered the evidence relating to clinical pregnancy to be of low quality for all dose comparisons, and moderate quality for comparisons of ORT algorithms versus dosing without ORT; limitations were risk of bias and imprecision (as for live birth).

## Potential biases in the review process

We conducted a comprehensive search with the help of an experienced Trials Search Co-ordinator, as well as extensive manual searching in an effort to retrieve all eligible studies; however, it remains possible that we may not have identified unpublished studies. We did identify one study that was published only as a conference abstract (Tasker 2010). We were unable to construct a funnel plot to investigate the possible extent of publication bias because fewer than 10 studies were available for all comparisons.

Although we contacted authors for additional information, we could not obtain all of the requested information, either because some trial authors did not provide this, or because we did not receive any reply to our queries. This may have introduced bias due to the inclusion of trials with insufficient information. Furthermore, there remains the potential for study authors to provide inaccurate information and overly positive answers.

It was difficult to decide on the structure of this review at the protocol stage (without knowing the types of studies or comparisons that would be included), so we planned to analyse each combination of protocols separately (e.g. for direct dose comparison studies: 150 IU versus 300 IU separate from 150 IU versus 450 IU, and for ORT-algorithm studies: AMH-based individualisation versus bFSH-based individualisation, and AMH-based individualisation versus AFC-based individualisation, etc.). It would not be meaningful to pool trials investigating different combinations of biomarkers in their study arms (for example, pooling a study of AMH-based individualisation versus bFSH-based individualisation with a study of AFC-based individualisation versus bFSH-based individualisation). At the outset of the review, we planned to pool only direct dose comparison trials that compared the same doses against one another. However, while conducting the review we found that most of the included studies had compared different dose sets, which made pooling impossible and limited the utility of the data. Therefore, we made a post hoc decision to pool studies in which the comparator dose (lower dose for predicted low and normal responders, and higher dose for predicted high responders) was the same (only within the same predicted response group). For example, in poor responders we pooled a trial comparing 150 IU versus 300 IU with a trial comparing 150 versus 450 IU. This is a major deviation from the protocol. Further, and of minor consequence, we decided to display all outcome data within a predicted response category on the same forest plots to streamline the review and improve visibility of trial results. As a consequence, review comparisons appear as 'subgroup' results (corresponding to different dose combinations) and are reported in the 'Summary of findings' tables. This is not normally recommended. However, after considering many possible configurations, we settled on the scheme presented here as the clearest way to present this complex review.

We also made other more minor post hoc decisions during the review process. For example, we decided to use the Peto OR for analysing multiple pregnancy data, as this outcome occurred at lower rates than anticipated. Another post hoc and arbitrary decision was to avoid the interpretation of treatment effects when there were five or fewer events in both trial arms.

As described previously, this review included two types of studies (direct dose comparison and ORT-algorithm studies) that answer different questions regarding the utility of individualised FSH dosing, and some may view the inclusion of direct dose comparison studies as only indirectly addressing the review aims.

## Agreements and disagreements with other studies or reviews

Although there is a lot of literature regarding the predictive ability of ORT measures and the association between increased dose and outcomes, we are aware of only one similar review of randomized controlled trials (Van Tilborg 2016). This review was also a systematic review on the same topic and included seven studies. The discrepancy between the number of included studies in Van Tilborg 2016 and our review appears largely due to differences in trial eligibility; the eligibility criteria in the other review was randomized trials "in which ORTs were used to determine the gonadotrophin starting dose", which resulted in the exclusion of some trials that we considered eligible. For example, we included studies that measured at least one ORT on all participants and provided the descriptions (e.g. mean and SD), as long as the study population could be reasonably categorised into predicted low, normal, or high responders. In contrast, Van Tilborg 2016 excluded these trials (e.g. Cavagna 2006; Hoomans 2002; Out 2004; YongPYK 2003). Additionally, this earlier review performed the literature search across only three databases, which may be the reason that review authors did not identify Tasker 2010 (a conference abstract) or Lan 2013. Further, although the review aimed to include only randomized studies, it did include one quasi-randomised study that we excluded based on author correspondence explaining that participants were allocated to trial arms based on their patient number rather than a truly random sequence (Berkkanoglu 2010).

These review authors elected not to pool any of the studies due to clinical and methodological heterogeneity of the studies, which we also encountered and was the reason that we only pooled some subgroups of studies in this review. Lack of author correspondence also led to different risk of bias judgements between the reviews, as we were able to obtain further clarification. Additionally, some of the outcomes for included studies differed, as we made adjustments to outcomes that were not reported appropriately according to the intention-to-treat principle. For example, we adjusted for censoring of the number of oocytes collected, which was commonly reported with the denominator of number of women reaching oocyte retrieval rather than the denominator of number of women randomized. Despite these differences and the additional trials included in our more recent review, the conclusions are similar. Both reviews state that there appears to be some benefit from individualising on upstream outcomes such as rates of hyper-response and the number of oocytes retrieved, but information on the most important clinical outcomes is limited. In light of recently published studies, we go slightly further than this by pointing out that differences in birth rates according to use of ORT are probably no more than a few percentage points, but OHSS risk may be reduced by ORT algorithms. Both reviews stated that most of the evidence is of low quality.

A similar author team has also conducted another review focusing on different doses in women under the age of 39 but not considering any ORT (Sterrenburg 2011).

# AUTHORS' CONCLUSIONS

## Implications for practice

We did not find that tailoring the FSH dose in any particular ORT population (low, normal, high ORT), influenced rates of live birth/ongoing pregnancy but we could not rule out differences, due to sample size limitations. In predicted high responders, lower doses of FSH seemed to reduce the overall incidence of moderate and severe OHSS. Moderate-quality evidence suggests that ORT-based individualisation produces similar live birth/ongoing pregnancy rates to a policy of giving all women 150 IU. However, in all cases the confidence intervals are consistent with an increase or decrease in the rate of around five percentage points with ORT-based dosing (e.g. from 25% to 20% or 30%). Although small, a difference of this magnitude could be important to many women. Further, ORT algorithms reduced the incidence of OHSS compared to standard dosing of 150 IU, probably by facilitating dose reductions in women with a predicted high response. However, the size of the effect is unclear. The included studies were heterogeneous in design, which limited the interpretation of pooled estimates, and many of the included studies had a serious risk of bias.

Current evidence does not provide a clear justification for adjusting the standard dose of 150 IU in the case of poor or normal responders, especially as increased dose is generally associated with greater total FSH dose and therefore greater cost. However, a decreased dose in predicted high responders may reduce OHSS.

Additionally, there is extra cost associated with obtaining ORT measures and administering the algorithms (e.g. a measure of AFC reported as costing EUR 62.52 in Van Tilborg 2017).

The knowledge of individual ORT measures may be helpful for clinicians in identifying a woman's anticipated response, as the data suggest that a lower dose is beneficial in high responders for reducing the risk of OHSS, with no observed impact on the probability of clinical pregnancy or live birth. An individual ORT measure may also be helpful for counselling patients about their predicted response, in terms of the probability of cycle cancellation for poor or hyper-response, and the probability of achieving live birth. Evidence suggests that women with low AMH/AFC values are less likely to achieve pregnancy from IVF than women with normal or high AMH/AFC. This may be useful information for patients planning to embark on an IVF cycle, especially if the treatment is at their own expense; they may wish to factor in the probability of success, which can be better predicted with knowledge of their own ORT. However, currently there is no evidence to suggest that any dose adjustment based on this ORT is beneficial to the chance of conceiving, and given that increased FSH dose is associated with increased cost of IVF, there may even be harm associated with increased dosing in women with predicted low or normal response. Further, clinicians should be aware that effects

of stimulation strategies, whether based on ORT measures or not, on increasing the number of oocytes retrieved have little clinical relevance. We lack evidence that these upstream outcomes impact on important endpoints such as live birth, and improvements in these upstream outcomes do not necessarily offer any benefit to couples undergoing IVF.

## Implications for research

Analysis of IVF data from randomized trials is complicated by the multistage nature of the treatment. We found multiple instances where trialists had reported or analyzed their outcomes in such a way so as to undermine the validity of the result. A common error was to report outcomes in subgroups of participants for whom treatment had not failed outright in the earlier stages. An illustrative example relates to the outcome measure 'number of oocytes'. This is a key mechanistic parameter in relation to this intervention, since the aim of tailoring ovarian stimulation to the individual is to reduce variation in the number of eggs obtained (i.e. to reduce poor or excessive egg yields). This is intended to safely maximise the chances of live birth. However, it was usual to report this outcome only in women who had a successful egg collection. If egg collection does not occur, however, it is usually because the IVF cycle has been cancelled for futility, and sometimes for hyper-response. Consequently, this approach will more often select out poor responders, exaggerating the expected egg yield. The exaggeration will be greater the higher the number of cancelled cycles, so there is scope for this approach to greatly overstate the benefits of treatment in women with low ovarian reserve. Further, if an individualised dosing strategy leads to a greater reduction in cycle cancellations in one group compared to the other, this would not be reflected in an outcome that only includes women reaching oocyte retrieval, and the subsequent result would be misleading. We recalculated the mean oocyte numbers in all women randomized by considering women with cancelled cycles to have zero eggs, and making an ad hoc inflation to the reported standard deviation. However, future researchers should be aware of the risks of selection bias to the internal validity of their trials. It is not an exaggeration to suggest that these incorrect analyses may turn results on their head.

Researchers should also be aware of the improvements in live birth rate that can realistically be detected in trials without rather large sample sizes. Even the larger studies included in the review were not powered to detect anything other than quite dramatic treatment effects (in the region of improvements in birth rates of 15% or higher). If studies are only powered to detect clear game-changers, then meaningfully better treatments will go overlooked. This is likely to represent an impediment to incremental progress. This is unfortunately the nature of randomized trials with binary primary outcomes such as live birth, the most important outcome, which makes these trials much more expensive and time-consuming to conduct. To illustrate: in order to detect an improvement in birth rate of 7% at a 5% significance threshold, sample sizes of around 2184 (at 80% power) or 2924 (at 90% power) would be required. These numbers far exceed those found in any study in the present review. Because studies were so heterogeneous, we were unable to convincingly overcome the limitation of study sizes. Although we did pool similar studies in a post hoc manner in order to increase sample size, the resulting estimates were still rather imprecise. This is a difficult problem, and if brute force (for example, large international collaboration projects) is not an option, then alternative (or complementary) strategies include methods to reduce biases in analysis of routinely collected data and methods capable of granting insight into the mechanisms of action of IVF (how does ovarian stimulation affect embryo implantation? How does it affect the chances that an implanted embryo will be sustained to term? And so on). A vast body of literature on casual inference methods exists, including methods for RCTs of complex interventions (Emsley 2009). These have not yet permeated the field of subfertility research, however.

The review provides very little direct information on safety of the included interventions, and large, carefully designed observational studies are likely to be needed to shed some light on the matter. We note that reanalysis of the trials included in this review (for example, using individual participant data) cannot overcome this problem; the information is not contained within the trial data. We also note that, while it is universally recognised as unacceptable, there is little agreement on how to define severe OHSS. We would suggest that this should be made a priority in subfertility research. Further, there was heterogeneity in the reporting of review outcomes, such as the definitions regarding the event of poor, normal or hyper-response. For example, one study defined a normal response as the retrieval of 5 to 12 oocytes (Magnusson 2017), while another defined it as 8 to 12 oocytes (Lan 2013).

Our 'Risk of bias' assessments have highlighted some recurrent design limitations in trials in this field. We would urge researchers to implement blinding in trials of dose individualisation, since there are many opportunities for clinician expectations to influence treatment decisions such as cycle cancellation and outcomes such as number of oocytes retrieved. We note that, as long as the initial dose allocation is concealed from the care team, it is not necessary to prohibit monitoring of the stimulation phase or even dose adjustment. It is legitimate - and not a source of bias - to make changes on the basis of response to treatment if blinding to treatment allocation is in place.

Finally, a recent multilevel modelling study attempted to quantify the extent to which variation in ovarian response could be anticipated and controlled for on the basis of current knowledge (Rustamov 2017). Even after taking into account many patient and treatment factors, including ORT and FSH dose, the authors found that an individual woman's response to stimulation was highly variable across multiple IVF cycles. This difficulty in predicting women's responses to treatment on the basis of their char-

acteristics presents an obstacle to individualised IVF. The same study also identified substantial variation in oocyte yield according to the surgeon performing the oocyte pickup. This suggests greater standardisation of some aspects of treatment as a possible alternative (or complement) to ORT-based individualisation, for the purpose of reducing variation in response. Future research should consider the importance of clinician variation in individualised treatment. Further, other strategies for increasing the probability of pregnancy from IVF, which may not include changes to FSH dosing, could be considered, for example the use of freeze-all strategy and modifications to IVF protocols, which have been demonstrated to affect outcomes in women with predicted low response (Pandian 2010).

## ACKNOWLEDGEMENTS

We thank the Cochrane Gynaecology and Fertility Group. In particular, we are grateful to Marian Showell (Information Specialist) for her assistance in developing the search strategies. We also thank the trial authors responding to our queries.

## REFERENCES

### References to studies included in this review

**Allegra 2017** *{published and unpublished data}*
Allegra A, Marino A, Volpes A, Coffaro F, Scaglione P, Gullo A, et al. A randomized controlled trial investigating the use of a predictive nomogram for the selection of the FSH starting dose in IVF/ICSI cycles. *Reproductive Biomedicine Online* 2017;**34**(4):429–38.

**Arce 2014** *{published data only}*
Arce JC, Andersen AN, Fernandez-Sanchez M, Visnova H, Bosch E, Garcia-Velasco JA, et al. Ovarian response to recombinant human follicle-stimulating hormone: a randomized, antimullerian hormone-stratified, dose response trial in women undergoing in vitro fertilization/ intracytoplasmic sperm injection. *Fertility and Sterility* 2014;**102**(6):1633–40.

**Bastu 2016** *{published and unpublished data}*
Bastu E, Buyru F, Ozsurmeli M, Demiral I, Dogan M, Yeh J. A randomized, single-blind, prospective trial comparing three different gonadotropin doses with or without addition of letrozole during ovulation stimulation in patients with poor ovarian response. *European Journal of Obstetrics & Gynecology and Reproductive Biology* 2016;**203**:30–4.

**Cavagna 2006** *{published data only}*
Cavagna M, Dzik A, Freitas GC, Soares JB, De Pawn K, Sales ALM, et al. Comparison of 150 IU and 225 IU of follitropin-beta in a fixed-dose regimen for ovarian stimulation using a depot formulation of GnRH agonist: a prospective randomized clinical trial. *Jornal Brasileiro de Reproducao Assistida* 2006;**10**(2):21–4.

**Harrison 2001** *{published data only}*
Harrison RF, Jacob S, Spillane H, Mallon E, Hennelly B. A prospective randomized clinical trial of differing starter doses of recombinant follicle-stimulating hormone (follitropin-$\beta$) for first time in vitro fertilization and intracytoplasmic sperm injection treatment cycles. *Fertility and Sterility* 2001;**75**(1):23–31.

**Hoomans 2002** *{published data only}*
Hoomans EHM, Mulder BB, on behalf of the Asian Puregon Study Group. A group-comparative, randomized, double-blind comparison of the efficacy and efficiency of two fixed daily dose regimens (100- and 200-IU) of recombinant follicle stimulating hormone (rFSH, Puregon) in Asian women undergoing ovarian stimulation for IVF/ ICSI. *Journal of Assisted Reproduction and Genetics* 2002;**19** (10):470–6.

**Jayaprakasan 2010** *{published and unpublished data}*
Jayaprakasan K, Hopkisson J, Campbell B, Johnson I, Thornton J, Raine-Fenning N. A randomised controlled trial of 300 versus225 IU recombinant FSH for ovarian stimulation in predicted normal responders by antral follicle count. *British Journal of Obstetrics and Gynaecology* 2010; **117**:853–62.

**Klinkert 2005** *{published and unpublished data}*
Klinkert ER, Broekmans FJM, Looman CWN, Habbema JDF, te Velde ER. Expected poor responders on the basis of an antral follicle count do not benefit from a higher starting dose of gonadotrophins in IVF treatment: a randomized controlled trial. *Human Reproduction* 2005;**20**(3):611–5.

**Lan 2013** *{published and unpublished data}*
Lan VTN, Linh NK, Tuong HM, Wong PC, Howles CM. Anti-Mullerian hormone versus antral follicle count for defining the starting dose of FSH. *Reproductive BioMedicine Online* 2013;**27**(4):390–9.

**Lefebrve 2015** *{published and unpublished data}*
Lefebvre J, Antaki R, Kadoch IJ, Dean NL, Sylvestre C, Bissonnette F, et al. 450 IU versus 600 IU gonadotropin for controlled ovarian stimulation in poor responders: a randomized controlled trial. *Fertility and Sterility* 2015;**104** (6):1419–25.

**Magnusson 2017** *{published and unpublished data}*
Magnusson Å, Nilsson L, Oleröd G, Thurin-Kjellberg1 A, Bergh C. The addition of anti-Müllerian hormone in an algorithm for individualized hormone dosage did not improve the prediction of ovarian response-a randomized, controlled trial. *Human Reproduction* 2017;**32**(4):811–9.

**Nyboe Andersen 2017** *{published and unpublished data}*
Nyboe Andersen A, Nelson SM, Fauser BCJM, García-Velasco JA, Klein BM, Arce JC, ESTHER-1 study group.

Individualized versus conventional ovarian stimulation for in vitro fertilization: a multicenter, randomized, controlled, assessor-blinded, phase3 noninferiority trial. *Fertility and Sterility* 2017;**107**(2):387–96.

**Olivennes 2015** ***{published data only (unpublished sought but not used)}***
Olivennes F, Trew G, Borini A, Broekmans F, Arriagada P, Warne DW, Howles CM. Randomized, controlled, open-label, noninferiority study of the CONSORT algorithm for individualized dosing of follitropin alfa. *Reproductive BioMedicine Online* 2015;**30**:248–57.

**Oudshoorn 2017** ***{published data only}***
Oudshoorn SC, van Tilborg TC, Eijkemans MJC, Oosterhuis GJE, Friederich J, van Hooff HA, et al. Individualized versus standard FSH dosing in women starting IVF/ICSI: an RCT. Part 2: the predicted hyper responder. *Human Reproduction* 2017;**32**(12):2506–14.

**Out 2004** ***{published data only}***
Out HJ, Rutherford A, Fleming R, Tay CCK, Trew G, Ledger W, et al. A randomized, double-blind, multicentre clinical trial comparing starting doses of 150 and 200 IU of recombinant FSH in women treated with the GnRH antagonist ganirelix for assisted reproduction. *Human Reproduction* 2004;**19**(1):90–5.

**Popovic-Todorovic 2003** ***{published and unpublished data}***
Popovic-Todorovic B, Loft A, Bredkjñer HE, Bangsbùll S, Nielsen IK, Nyboe Andersen A. A prospective randomized clinical trial comparing an individual dose of recombinant FSH based on predictive factors versus a 'standard' dose of 150 IU/day in 'standard' patients undergoing IVF/ICSI treatment. *Human Reproduction* 2003;**18**(11):2275–82.

**Tan 2005** ***{published data only}***
Tan SL, Child TJ, Cheung AP, Fluker MR, Yuzpe A, Casper R, et al. A randomized, double-blind, multicenter study comparing a starting dose of 100 IU or 200 IU of recombinant follicle stimulating hormone (Puregon R) in women undergoing controlled ovarian hyperstimulation for IVF treatment. *Journal of Assisted Reproduction and Genetics* 2005;**22**(2):81–8.

**Tasker 2010** ***{published and unpublished data}***
Tasker F, Hamoda H, Wilner H, Grace J, Khalaf Y. A randomised controlled trial comparing cycle outcome of individualised versus standard dosage of recombinant FSH with IVF/ICSI treatment. *Human Reproduction* 2010;**25** (Suppl 1):i170–210.

**Van Tilborg 2017** ***{published data only}***
Van Tilborg TC, Torrance HL, Oudshoorn SC, Eijkemans MJC, Koks CAM, Verhoeve HRV, et al. Individualized versus standard FSH dosing in women starting IVF/ICSI: anRCT. Part 1: the predicted poor responder. *Human Reproduction* 2017;**32**(12):2496–2505.

**YongPYK 2003** ***{published data only}***
Yong PYK, Brett S, Baird DT, Thong KJ. A prospective randomized clinical trial comparing 150 IU and 225 IU of recombinant follicle-stimulating hormone(Gonal-F*) in a fixed-dose regimen for controlled ovarian stimulation in in vitro fertilization treatment. *Fertility and Sterility* 2003;**79** (2):308–15.

## References to studies excluded from this review

**Berkkanoglu 2010** ***{published data only}***
Berkkanoglu M, Ozgur K. What is the optimum maximal gonadotropin dosage used in microdose flare-up cycles in poor responders?. *Fertility and Sterility* 2010;**94**(2):662–5.

**Camier 1999** ***{published data only}***
Camier B. A multicentre, prospective, randomised study to compare low dose protocol versus conventional administration of Rh-FSH (GONAL-F®) in normo-responders undergoing ART. *Congress of controversies in Obstestrics, Gynaecology and Infertility* 1999;**1**:7.

**Cavagna 2009** ***{published data only}***
Cavagna M, Dzik A, Freitas GC, Soares JB, Donadio N, Cavagna F, et al. A prospective and randomized trial comparing 225IU and 300 IU follitropin-alpha; in a fixed-dose regimen for controlled ovarian stimulation. *Jornal Brasileiro de Reproducao Assistida* 2009;**24**(2):66–70.

**DeJong 2000** ***{published data only}***
de Jong D, Macklon NS, Fauser BCJM. A pilot study involving minimal ovarian stimulation for in vitro fertilization: extending the "follicle-stimulating hormone window" combined with the gonadotropin-releasing hormone antagonist cetrorelix. *Fertility and Sterility* 2000; **73**(5):1051–4.

**Fluker 2000** ***{published data only}***
Fluker MR, Copeland JE, Yuzpe AA. A prospective randomized clinical trial of four fixed starting doses of recombinant FSH in women undergoing ovarian stimulation for IVF. *Human Reproduction* 2000;**15**:26.

**Latin-American Puregon IVF Study Group 2001** ***{published data only}***
Latin-American Puregon IVF Study Group. A double-blind clinical trial comparing a fixed daily dose of 150 and 250 IU of recombinant follicle-stimulating hormone in women undergoing in vitro fertilization. *Fertility and Sterility* 2001; **76**(5):950–6.

**NCT02915900** ***{unpublished data only}***
NCT02915900. Optimising FSH dosage during in vitro fertilization (IVF) [Method for Determining Optimal FSH Dosage During in Vitro Fertilization (IVF)]. clinicaltrials.gov/ct2/show/NCT02915900 First received 27 September 2016.

**Out 1999** ***{published data only}***
Out HJ, Lindenberg S, Mikkelsen AL, Eldar-Geva T, Healy DL, Leader A, et al. A prospective, randomized, double-blind clinical trial to study the efficacy and efficiency of a fixed dose of recombinant follicle stimulating hormone (Puregon) in women undergoing ovarian stimulation. *Human Reproduction* 1999;**14**(3):622–7.

**Out 2000** ***{published data only}***
Out HJ, Braat DDM, Lintsen BME, Gurgan T, Bukulmez O, Gökmen O, et al. Increasing the daily dose of

recombinant follicle stimulating hormone (Puregon) does not compensate for the age-related decline in retrievable oocytes are ovarian stimulation. *2000* Human Reproduction;**15**(1):29–35.

**Out 2001** ***{published data only}***
Out HJ, David I, Ron-El R, Friedler S, Shalev E, Geslevich J, et al. A randomized, double-blind clinical trial using fixed daily doses of 100 IU or 200 IU of recombinant FSH in ICSI cycles. *Human Reproduction* 2001;**16**(6):1104–9.

**Pruksananonda 2004** ***{published data only}***
Pruksananonda K, Suwajanakorn S, Sereepapong W, Virutamasen P. Comparison of two different fixed doses of follitropin-$\beta$ in controlled ovarian hyperstimulation: a prospective randomized, double blind clinical trial. *Journal of the Medical Association of Thailand* 2004;**87**(10):1151–5.

**Simberg 2000** ***{published data only}***
Simberg N, Rova K, Gudmundsson J, Lundkvist Ö, Lundqvist M, Bekuretsion M, et al. A randomized study comparing Puregon 100 IU or 150 IU as starting dose in IVF treatment. *Human Reproduction* 2000;**Abstract book 1**:125.

**Tsagareishvili 2005** ***{published data only}***
Tsagareishvili G, Khonelidze N, Lazarev A. Results of application of high daily doses of recombinant follicle stimulating hormone in different age groups of women in IVF program. Georgian Medical News. 2005; Vol. 6:11.

**Wikland 2001** ***{published data only}***
Wikland M, Bergh C, Borg K, Hillensjö T, Howles CM, Knutsson A, et al. A prospective, randomized comparison of two starting doses of recombinant FSH in combination with cetrorelix in women undergoing ovarian stimulation for IVF/ICSI. *Human Reproduction* 2001;**16**(8):1676–81.

## References to studies awaiting assessment

**NCT02309671** ***{published data only}***
NCT02309671. A Dose-response Trial Using FE 999049 in Japanese Women Undergoing in Vitro Fertilisation (IVF) / Intracytoplasmic Sperm Injection (ICSI) Treatment. https://clinicaltrials.gov/ct2/show/NCT02309671 First received 5th December 2014.

## References to ongoing studies

**CTRI/2016/10/007367** ***{unpublished data only}***
CTRI/2016/10/007367. Phase III Study to Evaluate the Efficacy and Safety of Recombinant Human FSH of Cadila Healthcare Limited, India as compared to Gonal-F Administered Subcutaneously in Female Patients Undergoing Assisted Reproductive Technology. http://ctri.nic.in/Clinicaltrials/pmaindet2.php?trialid=16431&EncHid=&userName=CTRI/2016/10/007367 First received 14th October 2016.

**EUCT2012-004969-40** ***{unpublished data only}***
EUCT2012-004969-40. An AMH based individualised controlled ovarian stimulation regiment using Corifollitrophin Alphs or graded doses of rFSH versus a standard protocol. A randomised controlled trial. www.clinicaltrialsregister.eu/ctr-search/trial/2012-004969-40/DK First received 23rd November 2012.

**NCT01794208** ***{unpublished data only}***
NCT01794208. Efficacy and Safety of FSH-GEX™ in Comparison With 150 IU Gonal-f®. https://clinicaltrials.gov/ct2/show/NCT01794208 First received 18th February 2013.

**NCT02430740** ***{unpublished data only}***
NCT02430740. Tailored Ovarian Stimulation Based on BMI, AMH, AFC. https://clinicaltrials.gov/ct2/show/NCT02430740 First received 30th April 2015.

**NCT02739269** ***{published data only}***
NCT02739269. Antimullerian Hormone Versus Antral Follicle Count for Determination of Gonadotrophin Dosing in IVF. https://clinicaltrials.gov/ct2/show/NCT02739269 First received 15th April 2016.

**Singh 2015** ***{published data only}***
Singh M, Singh R, Jindal A, Jindal PC. A prospective randomised controlled trial (RCT) on the role of AMH tailored stimulation protocols (Agonist or Antagonist), in improving IVF outcome in previous failed cycles. Human Reproduction. 2015; Vol. 30:i419-20, P-699.

## Additional references

**Alam 2017 [pers comm]**
Alam V. Query on trial: CONSORT vs 150 for FSH dosing. Lensen S 31st May 2017.

**Aljawoan 2012**
Aljawoan FY, Hunt LP, Gordon UD. Prediction of ovarian hyperstimulation syndrome in coasted patients in an IVF/ICSI program. *Journal of Human Reproductive Sciences* 2012;**5**(1):32–6.

**Andersen 2006**
Andersen AN, Devroey P, Arce JC. Clinical outcome following stimulation with highly purified hMG or recombinant FSH in patients undergoing IVF: a randomized assessor-blind controlled trial. *Human Reproduction* 2006; **21**(12):3217–27.

**Arce 2013**
Arce JC, La Marca A, Mirner Klein B, Nyboe Andersen A, Fleming R. Antimüllerian hormone in gonadotropin releasing-hormone antagonist cycles: prediction of ovarian response and cumulative treatment outcome in good-prognosis patients. *Fertility and sterility* 2013;**99**(6): 1644–53.

**ASRM 2016**
Practice Committee of the American Society for Reproductive Medicine. Prevention and treatment of moderate and severe ovarian hyperstimulation syndrome: a guideline. *Fertility and Sterility* 2016;**106**:1634–47.

**Bastu 2017 [pers comm]**
Bastu E. Query on trial: 450 vs 300 IU FSH. Eur J Obstet Gynecol Reprod Biol 2016. Lensen S 15 May 2017.

**Berkkanoglu 2017 [pers comm]**
Berkkanoglu M. Query on trial: 600 vs 450 vs 300 IU FSH. Fertility and sterility 2010. Lensen S 4th May 2017.

**Bourgain 2003**
Bourgain C, Devroey P. The endometrium in stimulated cycles for IVF. *Human Reproduction Update* 2003;**9**:515–22.

**Broekmans 2006**
Broekmans FJ, Kwee J, Hendriks DJ, Mol BW, Lambalk CB. A systematic review of tests predicting ovarian reserve and IVF outcome. *Human Reproduction Update* 2006;**12**: 685–718.

**Broekmans 2010**
Broekmans FJ, de Ziegler D, Howles CM, Gougeon A, Trew G, Olivennes F. The antral follicle count: practical recommendations for better standardization. *Fertility and Sterility* 2010;**94**(3):1044–51.

**Broekmans 2017 [pers comm]**
Broekmans F. Query on trial: Klinkert 2005. Lensen S 20 May 2017.

**Broer 2013a**
Broer SL, Dólleman M, van Disseldorp J, Broeze KA, Opmeer BC, Bossuyt PM, et al. Prediction of an excessive response in in vitro fertilization from patient characteristics and ovarian reserve tests and comparison in subgroups: an individual patient data meta-analysis. *Fertility and Sterility* 2013;**100**(2):420–9.

**Broer 2013b**
Broer SL, van Disseldorp J, Broeze KA, Dolleman M, Opmeer BC, Bossuyt P, et al. on behalf of the IMPORT study group. Added value of ovarian reserve testing on patient characteristics in the prediction of ovarian response and ongoing pregnancy: an individual patient data approach. *Human Reproduction Update* 2013;**19**(1):26–36.

**Broer 2014**
Broer SL, Broekmans FJM, Laven JSE, Fauser BCJM. Anti-Müllerian hormone: ovarian reserve testing and its potential clinical implications. *Human Reproduction Update* 2014;**20** (5):688–701.

**Calhaz-Jorge 2016**
Calhaz-Jorge C, De Geyter C, Kupka MS, de Mouzon, Erb K, Mocanu E, et al. Assisted reproductive technology in Europe, 2012: results generated from European registers: results generated from European registers by ESHRE. *Human Reproduction* 2016;**31**(8):1638–52.

**Dyer 2016**
Dyer S, Chambers GM, De Mouzon J, Nygren KG, Zegers-Hochschild F, Mansour R, et al. International committee for monitoring assisted reproductive technologies world report: Assistedreproductive technology 2008, 2009 and 2010. *Human Reproduction* 2016;**31**(7):1588–1690.

**Emsley 2009**
Emsley R, Dunn G, White IR. Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical Methods in Medical Research* 2009;**19**(3):237–70.

**Fauser 2017**
Fauser BCJM. Patient-tailored ovarian stimulation for in vitro fertilization. *Fertility and Sterility* 2017;**108**(4): 585–91.

**Ferraretti 2011**
Ferraretti A, La Marca A, Fauser BCJM, Tarlatzis B, Nargund G, Gianaroli L. ESHRE consensus on the definition of 'poor response' to ovarian stimulation for in vitro fertilization: the Bologna criteria. *Human Reproduction* 2011;**26**(7):1616–24.

**GRADEpro GDT 2014 [Computer program]**
GRADE Working Group, McMaster University. GRADEpro GDT. Version accessed 1 June 2017. Hamilton (ON): GRADE Working Group, McMaster University, 2014.

**Gunby 2010**
Gunby J, Bissonnette F, Librach C, Cowan L. Assisted reproductive technologies (ART) in Canada: 2006 results from the Canadian ART Register. *Fertility and Sterility* 2010;**93**:2189–201.

**Hamoda 2017 [pers comm]**
Hamoda H. Query on trial: individualised versus standard dosage of recombinant FSH. Lensen S 27th March 2017.

**Harbin Consensus Workshop Group 2014**
Harbin Consensus Workshop Group. Improving the Reporting of Clinical Trials of Infertility Treatments (IMPRINT): modifying the CONSORT statement. *Human Reproduction* 2014;**29**(10):2075–82.

**Harris 2016**
Harris K, Fitzgerald O, Paul R C, Macaldowie A, Lee E, Chambers GM. Assisted reproductive technology in Australia and New Zealand 2014. Sydney: National Perinatal Epidemiology and Statistics Unit, the University of New South Wales, 2016.

**Helmgaard 2017 [pers comm]**
Helmgaard L. Query on trial: Arce 2017. Lensen S 7 July 2017.

**Helmgaard 2017b [pers comm]**
Helmgaard L. n/a. Lensen S 7th July 2017.

**Higgins 2003**
Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**: 557–60.

**Higgins 2008**
Higgins JP, White IR, Anzures-Cabera J. Meta-analysis of skewed data: combining results reported on log-transformed or raw scales. *Statistics in Medicine* 2008;**27**:6072–92.

**Higgins 2011**
Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org. The Cochrane Collaboration.

**Jayaprakasan 2017 [pers comm]**
Jayaprakasan K. Query on trial: 225 v 300 IU FSH. 2010. Lensen S 30 May 2017.

**Kawwass 2015**
Kawwass JF, Kissin DM, Kulkarni AD, Creanga AA, Session DR, Callaghan WM, et al. Safety of assisted reproductive technology in the United States, 2000-2011. *JAMA* 2015; **313**(1):88.

**Kolibianakis 2002**
Kolibianakis E, Bourgain C, Albano C, Osmanagaoglu K, Smitz J, Van Steirteghem A, et al. Effect of ovarian stimulation with recombinant follicle-stimulating hormone, gonadotropin releasing hormone antagonists, and human chorionic gonadotropin on endometrial maturation on the day of oocyte pick-up. *Fertility and Sterility* 2002;**78**: 1025–9.

**Kupka 2014**
Kupka MS, Ferraretti AP, de Mouzon J, Erb K, D'Hooghe T, Castilla JA, et al. European IVF-Monitoring Consortium, for the European Society of Human Reproduction and Embryology (ESHRE). Assisted reproductive technology in Europe, 2010: results generated from European registers by ESHRE. *Human Reproduction* 2014;**29**(10):2099–113.

**La Marca 2012**
La Marca A, Papeleo E, Grisendi V, Argento C, Giulini S, Volpe A. Development of a nomogram based on markers of ovarian reserve for the individualisation of the follicle-stimulating hormone starting dose in in vitro fertilisation cycles. *BJOG* 2012;**119**(10):1171–9.

**La Marca 2014**
La Marca A, Sunkara SK. Individualization of controlled ovarian stimulation in IVF using ovarian reserve markers: from theory to practice. *Human Reproduction Update* 2014; **20**(1):124–40.

**La Marca 2017**
La Marca A, Minasi MG, Sighinolfi G, Greco P, Argento C, Grisendi V, et al. Female age, serum antimüllerian hormone level, and number of oocytes affect the rate and number of euploid blastocysts in in vitro fertilization/intracytoplasmic sperm injection cycles. *Fertility and Sterility* 2017;**108**(5): 777–83.

**La Marca 2017 [pers comm]**
La Marca A. Query on trial: Nomogram for FSH dosing 2017. Lensen S 24 March 2017.

**Labarta 2017**
Labarta E, Bosch E, Mercader A, Alamá P, Mateu E, Pellicer A. A higher ovarian response after stimulation for IVF is related to a higher number of euploid embryos. *BioMed Research International* 2017;**2017**:5637923.

**Lan 2017 [pers comm]**
Lan TN. Query on trial: AMH vs AFC trial for FSH dosing. Lensen S 18 May 2017.

**Lefebvre 2017 [pers comm]**
Lefebvre J. Query on trial: 450 vs 600 IU FSH. Lefebvre 2015. Lensen S 15th May 2017.

**Lekamge 2008**
Lekamge DN, Lane M, Gilchrist RB, Tremellen KP. Increased gonadotrophin stimulation does not improve IVF outcomes in patients with predicted poor ovarian reserve. *Journal of Assisted Reproduction and Genetics* 2008;**25**(11-12):515–21.

**Magnusson 2017 [pers comm]**
Magnusson A. Query on trial: Use of AMH in an algorithm for IVF. HR 2017. Lensen S 17th May 2017.

**McGowan 1985**
McGowan MR, Braithwaite M, Jochle W, Maplecoft RJ. Superovulation of beef heifers with Pergonal (HMG): a dose response trial. *Theriogenology* 1985;**24**(2):173–84.

**O'Brien 1979**
O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979;**35**(3):549–57.

**Pandian 2010**
Pandian Z, McTavish AR, Aucott L, Hamilton MPR, Bhattacharya S. Interventions for 'poor responders' to controlled ovarian hyperstimulation (COH) in in-vitro fertilisation (IVF). *Cochrane Database of Systematic Reviews* 2010, Issue 1. [DOI: 10.1002/14651858.CD004379.pub3

**Popovic-Todorovic 2017 [pers comm]**
Popovic-Todorovic B. Query on trial 150 vs 100-250 FSH for IVF. Lensen S 31st May 2017.

**Roberts 2005**
Roberts R, Iatropoulou A, Ciantar D, Stark J, Becker DL, Franks S, et al. Follicle-stimulating hormone affects metaphase I chromosome alignment and increases aneuploidy in mouse oocytes matured in vitro. *Biology of Reproduction* 2005;**72**(1):107–18.

**Rustamov 2017**
Rustamov O, Wilkinson J, La Marca A, Fitzgerald C, Roberts SA. How much variation in oocyte yield after controlled ovarian stimulation can be explained? A multilevel modelling study. *Human Reproduction Open* 2017;**2017**(3):1–10.

**Seifer 2002**
Seifer DB, MacLaughlin DT, Christian BP, Feng B, Shelden RM. Early follicular serum müllerian-inhibiting substance levels are associated with ovarian response during assisted reproductive technology cycles. *Fertility and Sterility* 2002; **77**(3):468–71.

**Sterrenburg 2011**
Sterrenburg MD, Veltman-Verhulst SM, Eijkemans MJC, Hughes EG, Macklon NS, Broekmans FJ, et al. Clinical outcomes in relation to the daily dose of recombinant follicle stimulating hormone for ovarian stimulation in in vitro fertilization in presumed normal responders younger than 39 years: a meta-analysis. *Human Reproduction Update* 2011;**17**(2):184–96.

**Steward 2014**
Steward RG, Lan L, Shah AA, Yeh JS, Price TM, Goldfarb JM, et al. Oocyte number as a predictor for ovarian hyperstimulation syndrome and live birth: an analysis of 256,381 in vitro fertilization cycles. *Fertility and Sterility* 2014;**101**(4):967–73.

**Sugano 1997**
Sugano M, Watanabe S. Use of highly purified porcine FSH preparation for superovulation in Japanese black cattle. *Journal of Veterinary Medical Science* 1997;**59**(3):223–5.

**Sunkara 2011**
Sunkara SK, Rittenberg V, Raine-Fenning N, Bhattacharya S, Zamora J, Coomarasamy A. Association between the number of eggs and live birth in IVF treatment: an analysis of 400 135 treatment cycles. *Human Reproduction* 2011;**26**(7):1768–74.

**Tajik 2013**
Tajik P, Zwinderman AH, Mol BW, Bossuyt PM. Trial designs for personalizing cancer care: a systematic review and classification. *Clinical Cancer Research* 2013;**19**(17):4578–88.

**Thoma 2013**
Thoma ME, McLain AC, Louis JF, King RB, Trumble AC, Sundaram R. Prevalence of infertility in the United States as estimated by the current duration approach and a traditional constructed approach. *Fertility and Sterility* 2013;**99**(5):1324–31.

**Thong 2017 [pers comm]**
Thong KJ. Query on trial: 150 v 225 IU FSH. Yong 2003. Lensen S 13th April 2017.

**Ting 2009**
Ting PY, Wen XY. The influence of FSH on the aneuploidy rate of human in vitro matured oocytes. *Fertility and Sterility* 2009;**92**(3):S14.

**Toftager 2017**
Toftager M, Bogstad J, Lossl K, Praetorius L, Zedeler A, Bryndorf T, et al. Cumulative live birthrates after one ART cycle including all subsequent frozen-thaw cycles in 1050 women: secondary outcome of an RCT comparing GnRH-antagonist and GnRH-agonist protocols. *Human Reproduction* 2017;**32**(3):556–67.

**Torrance 2017 [pers comm]**
Torrance H. Query on van Tilborg 2017. Lensen S 12th September 2017.

**Trew 2017 [pers comm]**
Trew G. Query on study: Out 2004. Lensen S 18th April 2017.

**Vail 2003**
Vail A, Gardener E. Common statistical errors in the design and analysis of subfertility trials. *Human Reproduction* 2003;**18**:1000–4.

**Van Rooij 2002**
Van Rooij IA, Broekmans FJ, te Velde ER, Fauser BC, Bancsi LF, de Jong FH, et al. Serum anti-Müllerian hormone levels: a novel measure of ovarian reserve. *Human Reproduction* 2002;**17**(12):3065–71.

**Van Tilborg 2016**
Van Tilborg TC, Broekmans FJM, Dolleman M, Eijkemans MJC, Mol BW, Laven JSE. Individualized follicle-stimulating hormone dosing and in vitro fertilization outcome in agonist downregulated cycles: a systematic review. *Acta Obstetricia et Gynecologica Scandinavica* 2016;**95**:1333–44.

**Visser 2006**
Visser JA, de Jong FH, Laven JS, Themmen AP. Anti-Müllerian hormone: a new marker for ovarian function. *Reproduction* 2006;**131**(1):1–9.

**Wong 2017**
Wong KM, van Wely M, Mol F, Repping S, Mastenbroek S. Fresh versus frozen embryo transfers in assisted reproduction. *Cochrane Database of Systematic Reviews* 2017, Issue 3. [DOI: 10.1002/14651858.CD011184.pub2

**Youssef 2016**
Youssef M, Mourad S. Volume expanders for the prevention of ovarian hyperstimulation syndrome. *Cochrane Database of Systematic Reviews* 2016, Issue 11. [DOI: 10.1002/14651858.CD001302

## References to other published versions of this review

**Lensen**
Lensen SF, Wilkinson J, Mol BWJ, La Marca A, Torrance H, Broekmans FJ. Individualised gonadotropin dose selection using markers of ovarian reserve for women undergoing IVF/ICSI. Individualised gonadotropin dose selection using markers of ovarian reserve for women undergoing IVF/ICSI. *Cochrane Database of Systematic Reviews* 2017, Issue 6. [DOI: 10.1002/14651858.CD012693

* *Indicates the major publication for the study*

# CHARACTERISTICS OF STUDIES

## Characteristics of included studies *[author-defined order]*

### Allegra 2017

| | |
|---|---|
| Methods | RCT, **ORT-algorithm study**, 194 randomized<br>Study arms: 2<br>Setting: Italy, Andros Day Surgery Clinic<br>Recruitment period: January 2011 to April 2015* |
| Participants | Age: eligibility criteria 18-40 years. Mean (SD) G1: 34.4 years (3.9), G2: 33.5 years (4.3)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>To be included participants had to have AMH concentrations between 1.0 and 4.0 ng/mL and basal FSH ≤ 15 IU/L<br>**AMH**: eligibility criteria 1-4 ng/ml. Mean (SD) G1: 17.9 pmol/L (6.4), G2: 17.1 pmol/L (7.2)(converted from ng/L to pmol/L). Assay: modified AMH Gen II ELISA<br>**AFC**: no eligibility criteria: mean (SD) G1: 10.8 (4.9), G2: 11.7 (5.5). AFC definition: follicles of 2-9 mm both ovaries*<br>**bFSH (IU/L)**: eligibility criteria ≤ 15 IU/L. Mean (SD) G1: 7.9 IU/L (5.3), G2: 7.1 IU/L (2.5)<br>**Criteria relating to prior IVF/ICSI response**: no<br>Additional inclusion criteria: first IVF cycle, BMI 18-25 kg/$m^2$, normal regular menstrual cycles, ranging from 25 to 33 days in length, normal thyroid-stimulating hormone (TSH) and prolactin concentrations, normal uterine cavity as assessed by hysteroscopy or sonohysterography or 3-dimensional ultrasound and presence of both ovaries<br>Additional exclusion criteria: irregular menstrual cycles, PCOS, severe endometriosis, previous ovarian surgery, presence of ovarian cysts, use of hormonal contraception in the previous 3 months, any known metabolic or endocrinological disease |
| Interventions | Group 1: **dose based only on age (150 IU ≤ 35 years,225 IU > 35 years)**<br>Group 2: **dose based on nomogram including age,bFSH,AMH** (see paper)<br>Protocol: long agonist<br>Dose titration: adjustments were permitted after the first scan (day 5/6) |
| Outcomes | Reported in paper: ongoing pregnancy, clinical pregnancy, number of oocytes retrieved, total dose of FSH, duration of FSH, cycle cancellations for poor and hyper-response, poor response, normal response, hyper-response, moderate and severe OHSS, had at least 1 transferable embryo<br>Obtained from author correspondence*: none<br>Not available: live birth, multiple pregnancy |
| Notes | Trial registration: NCT01816789, retrospective (after commencing recruitment)<br>Funding: "This research did not receive any specific grant from funding agencies in the public, commercial, or not for-profit sectors"<br>Conflict of interest: "The authors report no financial or commercial conflicts of interest"<br>Other presentation?<br>Email correspondence undertaken: yes, with A La Marca, a co-author of this trial and review author (La Marca 2017 [pers comm]) |

| | |
|---|---|
| | *indicates information obtained from email correspondence |

***Risk of bias***

| Bias | Authors' judgement | Support for judgement |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper stated, "randomly assigned to one of two treatment groups by giving them a code number from a randomization sequence (in order of enrolment). The randomization sequence was generated by a computer program software (PASW-17) using a simple randomization method." |
| Allocation concealment (selection bias) | Low risk | Paper describes third-party randomization: "to guarantee the concealment of allocation, a staff member, who was not involved in the study, was in possession of the randomization sequence; in this way, after receiving information from the physician recruiting the couples, the staff member followed the randomization sequence allocating each couple" |
| Blinding of participants and personnel (performance bias)<br>All outcomes | High risk | Authors confirmed no blinding used* |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | High risk | Authors confirmed no blinding used* |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | 3 women excluded postrandomisation, small numbers |
| Selective reporting (reporting bias) | High risk | Post hoc changes to primary outcome definition |
| Other bias | High risk | Stopping on the basis of O'Brien and Fleming 1979 rules means that the data available to us as systematic reviewers give a biased effect estimate |

**Arce 2014**

| | |
|---|---|
| Methods | RCT, **direct dose comparison study**, 265 randomized<br>Study arms: 6 arms (5 relevant and included in this review)<br>Setting: 7 centres in 4 countries (Belgium, Czech Republic, Denmark, and Spain)<br>Recruitment period: September 2011 to May 2018 |
| Participants | Population: data has been stratified into anticipated **NORMAL** responders and anticipated **HIGH** responders - presented separately in this review<br>Age: eligibility criteria: 18-37 years. Mean (SD) G1: 33.6 years (2.2), G2: 32.3 years (3.5), G3: 32.8 years (2.4), G4: 32.3 years (3.2), G5: 32.6 years (3.0)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>To be eligible participants had to have early follicular phase FSH serum concentration of 1-12 IU/L and AFC ≥ 6 and ≤ 25 for both ovaries combined; serum AMH concentration of 5.0-44.9 pmol/L (0.7-6.3 ng/mL)<br>Anticipated **NORMAL** responders<br>**AMH (pmol/L)**: eligibility criteria: 5.0-14.9 (stratification). Median (IQR) G1: 9 pmol/L (7-11), G2: 9 pmol/L (7-12), G3: 9 pmol/L (7-11), G4: 10 pmol/L (8-13), G5: 10 pmol/L (7-11). Assay: Beckman Coulter Gen 2 ELISA<br>**AFC**: mean (SD) G1: 11.6 (3.7), G2: 11.5 (2.9), G3: 11.6 (3.6), G4: 12.0 (3.2), G5: 13.3 (4.8). AFC definition: follicles of 2-10 mm in both ovaries<br>**bFSH (IU/L)**: median (IQR) G1: 6.1 IU/L (5.7-7.8), G2: 7.2 IU/L (5.4-8.2), G3: 7.9 IU/L (6.6-9.3), G4: 7.6 IU/L (6.5-8.5), G5: 8.0 IU/L (6.2-10.1)<br>Anticipated **HIGH** responders<br>**AMH (pmol/L)**: eligibility criteria: 15.0-44.9 (stratification). Median (IQR) G1: 23 pmol/L (17-29), G2: 26 pmol/L (19-29), G3: 22 pmol/L (19-29), G4: 25 pmol/L (21-34), G5: 26 pmol/L (19-31). Assay: Beckman Coulter Gen 2 ELISA a<br>**AFC**: mean (SD) G1: 15.4 (4.3), G2: 14.5 (5.3), G3: 15.2 (4.4), G4: 16.7 (4.1), G5: 15.1 (4.2). AFC definition: follicles of 2-10 mm in both ovaries<br>**bFSH (IU/L)**: median (IQR) G1: 6.6 IU/L (4.9-7.3), G2: 6.8 IU/L (5.8-7.3), G3: 6.2 IU/L (5.3-6.7), G4: 6.6 IU/L (5.3-7.6), G5: 6.9 IU/L (6.1-8.1)<br>**Criteria relating to prior IVF/ICSI response**: yes, women were excluded if they had: a poor ovarian response in a previous IVF/ICSI cycle using an average daily FSH dose ≥ 150 IU, defined as development of fewer than 4 follicles ≥ 15 mm or cycle cancellation due to limited follicular response; excessive ovarian response in a previous IVF/ICSI cycle using an average daily FSH dose < 225 IU, defined as > 25 oocytes retrieved or cycle cancellation due to excessive ovarian response, including risk of OHSS; severe OHSS in a previous IVF/ICSI cycle<br>Additional inclusion criteria: diagnosed with tubal infertility, unexplained infertility, infertility related to endometriosis stage I/II, or with partners diagnosed with male factor infertility, BMI 18.5-32.0 kg/m$^2$; infertility for at least 1 year before randomization; regular menstrual cycles of 24-35 days, presumed to be ovulatory; hysterosalpingography, hysteroscopy, or transvaginal ultrasound documenting a uterus consistent with expected normal function; transvaginal ultrasound documenting presence and adequate visualisation of both ovaries, without evidence of significant abnormality; willing to accept transfer of 1 blastocyst in the fresh cycle; and willing to accept transfer of 1 blastocyst in frozen embryo replacement cycles initiated within 6 months after randomization<br>Additional exclusion criteria: known PCOS associated with anovulation; known endometriosis stage III-IV; 3 or more stimulated cycles for IVF/ICSI; history of recurrent miscarriage; current or past (up to 1 year before randomization) abuse of alcohol or drugs; and intake of more than 14 units of alcohol per week during the past month or smoking more than 10 cigarettes per day within 3 months before randomization |

**Arce 2014** (*Continued*)

| | |
|---|---|
| Interventions | Group 1 dose/drug: **5.2 µg** (FE 999049, Ferring Pharmaceuticals)<br>Group 2 dose/drug: **6.9 µg** (FE 999049, Ferring Pharmaceuticals)<br>Group 3 dose/drug: **8.6 µg** (FE 999049, Ferring Pharmaceuticals)<br>Group 4 dose/drug: **10.3 µg** (FE 999049, Ferring Pharmaceuticals)<br>Group 5 dose/drug: **12.1 µg** (FE 999049, Ferring Pharmaceuticals)<br>Protocol: antagonist<br>Dose titration: not permitted |
| Outcomes | Reported in paper: live birth, clinical pregnancy, number of oocytes, duration of FSH, moderate and severe OHSS, had at least 1 transferable embryo, poor response, normal response, hyper-response<br>Obtained from author correspondence*: multiple pregnancy, total dose of FSH, cycle cancellations for poor and hyper-response<br>Not available: ongoing pregnancy |
| Notes | Trial registration: NCT01426386<br>Funding: Ferring Pharmaceuticals<br>Conflict of interest: many investigators have received payments or grants from Industry<br>The results of this study were reported at American Society for Reproductive Medicine conference 2013 (Arce 2013)*<br>Email correspondence undertaken: yes, with corresponding author Joan-Carles Arce and his colleagues (Helmgaard 2017 [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| Bias | Authors' judgement | Support for judgement |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper states, "[a]n independent statistician in the Department of Biometrics at Ferring Pharmaceuticals generated the randomization list using SAS version 9.2 (SAS Institute)" |
| Allocation concealment (selection bias) | Low risk | Paper states, "[r]andomization was performed centrally through the electronic Case Report Form system by assigning the lowest randomization number available within stratum" Therefore allocation would be concealed until the moment of randomization |
| Blinding of participants and personnel (performance bias)<br>All outcomes | Unclear risk | Paper states, "[t]he trial was assessor-blinded, and all investigators, embryologists, central laboratory personnel, and sponsor staff involved in analyzing and interpreting data were kept blinded to treatment allocation throughout the trial. |

| | | |
|---|---|---|
| | | " However, the participants were not blind and may have disclosed their dose allocation to the staff, and it is unclear whether there were any safeguards to prevent this |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | Unclear risk | Paper states, "[t]he trial was assessor-blinded, and all investigators, embryologists, central laboratory personnel, and sponsor staff involved in analyzing and interpreting data were kept blinded to treatment allocation throughout the trial. " However, the participants were not blind and may have disclosed their dose allocation to the staff, and it is unclear whether there were any safeguards to prevent this |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | Only 1 participant dropped out for personal reasons |
| Selective reporting (reporting bias) | High risk | Live birth was not listed as an outcome on the trial registration website; however, trial authors confirmed that they always planned to capture live-birth and cumulative live birth data as it was considered a follow-up activity rather than an outcome. Additionally, total dose of FSH was listed as an outcome on the trial registration but not reported in the paper; however, this was provided by the authors upon request |
| Other bias | Low risk | - |

**Bastu 2016**

| | |
|---|---|
| Methods | RCT, **direct dose comparison study**, 62 randomized<br>Study arms: 2 (3 arms in paper, but 3rd arm not included in this review as participants administered letrozole)<br>Setting: Turkey, 1 centre: Istanbul University School of Medicine<br>Recruitment period: November 2014 - August 2015 |
| Participants | Population: anticipated **LOW** responders<br>Age: eligibility criteria: 18-42 years. Mean (SD) G1: 36.94 years (3.33), G2: 35.00 years (3.10)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>Participants met the Bologna criteria: at least 2 of the following 3 criteria had to be met: (advanced maternal age (40 years) and/or any other risk factor for poor ovarian response; previous history of poor ovarian response (retrieval of 3 oocytes during conventional stimulation protocol); and an abnormal ORT (AMH or AFC*)<br>**AMH (pmol/L)**: eligibility criteria: 'abnormal ORT' Abnormal AMH was defined as |

| | |
|---|---|
| | AMH < 7.8 pmol/L*. Mean (SD) G1: 3.9 pmol/L (2.5), G2: 5.1 pmol/L (2.0) (converted from ng/mL). Assay: AMH Gen II, Beckman Coulter, US*<br>**AFC**: eligibility criteria: 'abnormal ORT' Abnormal was defined as AFC < 7*. Mean (SD) G1: 3 (not provided), G2: 4 (not provided). AFC definition: follicles in both ovaries measuring 2-9 mm*<br>**bFSH (IU/L)**: no eligibility criteria: G1: 10.63 IU/L (3.95), G2: 11.01 IU/L (2.34)<br>**Criteria relating to prior IVF/ICSI response**: yes, previous poor response an eligibility criteria<br>Additional inclusion criteria aged between 18 and 42 years, regular menstrual cycles (menstrual cycles of 25-34 days), normal BMI of 19.3-28.9 kg/m$^2$, no metabolic or endocrine disorders, normal hormone panel, couples undergoing the ICSI cycle with ejaculated sperm, normal uterine documented by hysterosalpingography or hysteroscopy<br>Additional exclusion criteria: history of cytotoxic chemotherapy and/or radiotherapy, history of ovarian surgery such as oophorectomy or cystectomy, history of dehydroepiandrosterone (DHEA) and/or testosterone supplement use, women undergoing natural IVF cycle |
| Interventions | Group 1 dose/drug: **300 IU** (225 IU hMG; Menogon; Ferring and 225 IU rFSH; Gonal-F; Merck KGaA)<br>Group 2 dose/drug: **450 IU** (225 IU hMG; Menogon; Ferring and 225 IU rFSH; Gonal-F; Merck KGaA)<br>Protocol: antagonist<br>Dose titration: not permitted* |
| Outcomes | Reported in paper: ongoing pregnancy, clinical pregnancy, number of oocytes, total dose of FSH, duration of FSH<br>Obtained from author correspondence*: multiple pregnancy, moderate and severe OHSS, had at least 1 transferable embryo, cycle cancellations for poor and hyper-response<br>Not available: live birth, poor response, normal response, hyper-response |
| Notes | Trial registration: NCT02293668: registered prospectively<br>Funding: no funding*<br>Conflict of interest: none disclosed<br>Preliminary data was presented in the American Society of Reproductive Medicine (ASRM) Scientific Congress in 2015*<br>Email correspondence undertaken: yes with Ercan Bastu (Bastu 2017 [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| **Bias** | **Authors' judgement** | **Support for judgement** |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper states, "[t]he randomization list was a computer-generated sequence" |
| Allocation concealment (selection bias) | Low risk | Paper states, "[s]ealed envelopes were used for the randomization list"; however SNOSE envelopes confirmed following au- |

**Bastu 2016** (*Continued*)

| | | |
|---|---|---|
| | | thor correspondence "opaque envelopes that were numbered in sequence"* |
| Blinding of participants and personnel (performance bias)<br>All outcomes | Unclear risk | "The infertility specialist (E.B.) who was blinded observed follicular development using ultrasound and retrieved oocytes in all participating patients. The embryologist (S.B.) was also blinded to the assigned treatment protocol." However, the participants were not blind and may have disclosed their dose allocation to the staff, and it is unclear whether there were any safeguards to prevent this |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | Unclear risk | "The infertility specialist (E.B.) who was blinded observed follicular development using ultrasound and retrieved oocytes in all participating patients. The embryologist (S.B.) was also blinded to the assigned treatment protocol." However, the participants were not blind and may have disclosed their dose allocation to the staff, and it is unclear whether there were any safeguards to prevent this |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | Author correspondence confirmed there was no attrition. |
| Selective reporting (reporting bias) | Low risk | All outcomes in trial registration are reported |
| Other bias | Low risk | - |

**Cavagna 2006**

| | |
|---|---|
| Methods | RCT, **direct dose comparison study**, 76 randomized<br>Study arms: 2<br>Setting: Brazil, 1 centre: Unit of Assisted Reproduction of the Centro de Referencia da Saude da Mulher, Hospital Perola Byington<br>Recruitment period: not stated |
| Participants | Population: anticipated **NORMAL** responders<br>Age: eligibility criteria 18-35 years. Mean (SD) G1: 31.4 years (2.8), G2: 31.7 years (2.8)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>**bFSH (IU/L):** eligibility < 10 IU/L<br>**AMH** no eligibility criteria, not recorded<br>**AFC** no eligibility criteria, not recorded<br>**bFSH (IU/L)**: eligibility criteria < 10. Mean (SD) not provided |

**Cavagna 2006** (*Continued*)

| | |
|---|---|
| | **Criteria relating to prior IVF/ICSI response**: yes "previous ART cycle with poor response to stimulation"<br>Additional inclusion criteria: indicated for IVF/ICSI, normal menstrual cycle (range 24-35 days), BMI 19-29 kg/$m^2$<br>Additional exclusion criteria: endocrine abnormalities, systemic chronic disease |
| Interventions | Group 1 dose/drug: **150 IU** (Puregon, Organon)<br>Group 2 dose/drug: **200 IU** (Puregon, Organon)<br>Protocol: long agonist<br>Dose titration: not permitted |
| Outcomes | Reported in paper: clinical pregnancy, moderate and severe OHSS, cancellations for poor and hyper-response, total dose of FSH, had at least 1 transferable embryo, number of oocytes<br>Obtained from author correspondence: none<br>Not available: ongoing pregnancy, live birth, normal response, hyper-response, poor response, multiple pregnancy, duration of FSH |
| Notes | Trial registration: none<br>Funding: not stated<br>Conflict of interest: not stated<br>Email correspondence undertaken: attempted, but no response |

***Risk of bias***

| **Bias** | **Authors' judgement** | **Support for judgement** |
|---|---|---|
| Random sequence generation (selection bias) | Unclear risk | Stated as "randomized" only |
| Allocation concealment (selection bias) | Unclear risk | Stated as "randomized" only |
| Blinding of participants and personnel (performance bias)<br>All outcomes | High risk | No description of blinding, assume unblinded |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | High risk | No description of blinding, assume unblinded |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | Appears to be no attrition |
| Selective reporting (reporting bias) | High risk | Does not appear to be registered, protocol not available, and not reporting important outcomes such as live birth |
| Other bias | Low risk | - |

**Harrison 2001**

| | |
|---|---|
| Methods | RCT, **direct dose comparison study**, 297 included (unclear exact number randomized)<br>Study arms: 2 (this study has 4 arms, 2 are included as Harrison 2001a, and 2 included in Harrison 2001b)<br>Setting: Ireland, 1 centre: Rotunda Hospital<br>Recruitment period: 1 January to 31 December 1997 |
| Participants | Data is stratified in study based on both **NORMAL** and **LOW** responders<br>Population: anticipated **NORMAL** responders<br>Age: no eligibility criteria. Mean (SD) G1: 34.0 years (3.5), G2: 33.3 years (4.0)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>**AMH** no eligibility criteria, not recorded<br>**AFC** no eligibility criteria, not recorded<br>**bFSH (IU/L)**: eligibility criteria < 8.5 IU/L. Mean (SD) G1: 5.9 IU/L (1.3), G2: 6.0 IU/L (1.3)<br>Population: anticipated **LOW** responders<br>Age: no eligibility criteria. Mean (SD) G1: 35.0 years (3.9), G2: 36.2 years (3.6)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>**AMH** no eligibility criteria, not recorded<br>**AFC** no eligibility criteria, not recorded<br>**bFSH (IU/L):** eligibility criteria > 8.5 IU/L. Mean (SD) G1: 10 IU/L (1.5), G2: 10.2 IU/L (1.4)<br>**Criteria relating to prior IVF/ICSI response**: no<br>Additional inclusion criteria: undergoing first IVF/ICSI cycle<br>Additional exclusion criteria: - |
| Interventions | Group 1 dose/drug: **150 IU** (follitropin-beta Puregon)<br>Group 2 dose/drug: **200 IU** (follitropin-beta Puregon)<br>Group 3 dose/drug: **300 IU** (follitropin-beta Puregon)<br>Group 4 dose/drug: **400 IU** (follitropin-beta Puregon)<br>Protocol: long agonist<br>Dose titration: if the response to the starting dose of FSH was poor (fewer than 3 follicles, and E2 level of < 300 pmol/L) on day 5 of stimulation treatment the FSH dose was doubled (except for Group 2/400 IU dose, where a maximum of 600 IU was used) |
| Outcomes | Reported in paper: clinical pregnancy, cancellations for poor and hyper-response, poor response, total dose of FSH, duration of FSH, had at least 1 transferable embryo, number of oocytes<br>Obtained from author correspondence: none<br>Not available: ongoing pregnancy, live birth, normal response, hyper-response, moderate and severe OHSS, multiple pregnancy |
| Notes | Trial registration: not registered<br>Funding: Organon UK<br>Conflict of interest:<br>Email correspondence undertaken: co-authors of the trial informed us the corresponding author now deceased; study data not available |
| ***Risk of bias*** | |

**Harrison 2001** (*Continued*)

| Bias | Authors' judgement | Support for judgement |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper states randomization achieved with the "use of a computer-generated list provided by Organon Ltd" |
| Allocation concealment (selection bias) | Low risk | Paper states "starter dosages of follitropin-beta were randomised through the hospital pharmacy, and they were blinded to the clinicians" |
| Blinding of participants and personnel (performance bias)<br>All outcomes | High risk | Described as "open", therefore no blinding employed |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | Low risk | Described as "open", therefore no blinding employed; however, this study did not report OHSS, therefore this risk of bias is not relevant |
| Incomplete outcome data (attrition bias)<br>All outcomes | High risk | 19 women excluded for various reasons: wrong dose, other violations. It is unclear which groups the participants were excluded from, therefore it is impossible to perform ITT analysis |
| Selective reporting (reporting bias) | High risk | Study not registered, no protocol available, live birth not reported |
| Other bias | Low risk | - |

**Hoomans 2002**

| | |
|---|---|
| Methods | RCT, **direct dose comparison study**, 330 randomized<br>Study arms: 2<br>Setting: 9 study centres: 2 in Hong Kong, 3 each in India and Thailand, 1 in Singapore<br>Recruitment period: December 1997 to July 1999 |
| Participants | Population: anticipated **NORMAL** responders<br>Age: eligibility criteria: 18-39 years. Mean (SD) G1: 31.6 years (3.6), G2: 32.1 years (3.8)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>**AMH**: no eligibility criteria<br>**AFC**: no eligibility criteria<br>**bFSH (IU/L)**: no eligibility criteria: mean (SD) G1: 3.8 IU/L (SD not provided), 3.5 IU/L (SD not provided)<br>**Criteria relating to prior IVF/ICSI response**: yes, excluded if previous assisted reproduction in which fewer than 3 oocytes were retrieved or previous hospitalisation due to |

| | |
|---|---|
| | severe ovarian hyperstimulation syndrome<br>Additional inclusion criteria: normal ovulatory cycles with a mean length of 24-35 days, good physical and mental health, and BMI of 18-29 kg/$m^2$<br>Additional exclusion criteria: infertility caused by endocrine abnormalities such as hyperprolactinaemia, PCOS and absence of ovarian function; if they suffered from chronic cardiovascular, hepatic, renal, or pulmonary disease; had a history of (within 12 months) or currently indulged in abuse of alcohol or drugs; or had used investigational drugs within 3 months before screening |
| Interventions | Group 1 dose/drug: **100 IU** (Puregon, Organon)<br>Group 2 dose/drug: **200 IU** (Puregon, Organon)<br>Protocol: long agonist<br>Dose titration: presume no dose adjustment permitted as title is "Comparison of ... two fixed daily dose regimens" |
| Outcomes | Reported in paper: ongoing pregnancy, clinical pregnancy, moderate and severe OHSS, number of oocytes, total dose of FSH, duration of FSH, cancellations for poor and hyper-response, had at least 1 transferable embryo, multiple pregnancy<br>Obtained from author correspondence: none<br>Not available: live birth, poor response, normal response, hyper-response |
| Notes | Trial registration: none stated<br>Funding: Organon<br>Conflict of interest: none declared<br>Email correspondence undertaken: no, unable to contact any of the author team |

***Risk of bias***

| **Bias** | **Authors' judgement** | **Support for judgement** |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper states, "computer-generated randomization list using random numbers" |
| Allocation concealment (selection bias) | Unclear risk | No description |
| Blinding of participants and personnel (performance bias)<br>All outcomes | Low risk | Described as double-blind and the FSH "was supplied as lyophilized spheres (lyospheres) in ampoules containing 50- or 100-IU FSH in vivo bioactivity. For s.c. injection, lyospheres were reconstituted with 1mL of solvent" which suggests the medications would be indistinguishable, therefore likely all participants and personnel were blind |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | Low risk | Described as double-blind, and the FSH "was supplied as lyophilized spheres (lyospheres) in ampoules containing 50- or 100-IU FSH in vivo bioactivity. For s.c. in- |

**Hoomans 2002** (*Continued*)

| | | |
|---|---|---|
| | | jection, lyospheres were reconstituted with 1mL of solvent", which suggests the medications would be indistinguishable, therefore likely all participants and personnel were blind, and likely the assessor would also be blind |
| Incomplete outcome data (attrition bias) All outcomes | Low risk | Appears to be no study attrition |
| Selective reporting (reporting bias) | High risk | Trial not registered, no protocol available, not reporting important outcomes such as live birth |
| Other bias | High risk | The sample size was increased on the basis of interim analysis. There does not appear to have been any correction for this in the trial. More pertinently, this means that the summary data available for this review will represent a biased estimate of the treatment effect |

**Jayaprakasan 2010**

| | |
|---|---|
| Methods | RCT, **direct dose comparison study**, 135 randomized<br>Study arms: 2<br>Setting: UK, 1-centre: Nottingham University Research and Treatment Unit in Reproduction (NURTURE)<br>Recruitment period: September 2006 to March 2008 |
| Participants | Population: anticipated **NORMAL** responders<br>Age: eligibility criteria: < 39 years old. Mean (SD) G1: 34.2 years (3.5), 33.1 years (3.7)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>Eligible if bFSH level was below 12 IU/L, and if their total AFC was 8-21<br>**AMH (pmol/L)**: no eligibility criteria: mean (SD) G1: 9.3 pmol/L (5.0), G2: 10.0 pmol/L (5.7) - converted from ng/mL. Assay: DSL 9 Diagnostic system Lab*<br>**AFC**: eligibility criteria: 8-21. Mean (SD) G1: 14.1 ± 4.0, G2: 15.4 ± 3.9 (8-21). AFC definition: follicles in both ovaries measuring 2-10 mm<br>**bFSH (IU/L)**: eligibility criteria: < 12 IU/L. Mean (SD) G1: 7.3 IU/L (1.8), G2: 6.8 IU/L (1.8)<br>**Criteria relating to prior IVF/ICSI response**: no<br>Additional inclusion criteria: both ovaries were present, no past history of ovarian surgery, regular spontaneous menstrual cycle of 21-35 days, and BMI 20-35 kg/$m^2$<br>Additional exclusion criteria: presence of an ovarian cyst or a follicle measuring 20 mm or more in diameter, or other significant pelvic pathology, such as fibroids, a hydrosalpinx, an endometrioma, or a uterine anomaly. Participants with unilateral or bilateral polycystic ovaries, as defined by the Rotterdam criteria of the presence of 12 or more follicles measuring 2-9 mm in diameter, or an ovarian volume of more than 10 $cm^3$, were also excluded. Participants were also excluded if they were known to have PCOS or |

| | |
|---|---|
| | if their subfertility was related to, or associated with, any other recognised endocrine abnormalities, such as hyperprolactinaemia, thyroid dysfunction, and hyperandrogenism |
| Interventions | Group 1 dose/drug: **225 IU** (Gonal-F; Merck Serono)<br>Group 2 dose/drug: **300 IU** (Gonal-F; Merck Serono)<br>Protocol: long agonist<br>Dose titration: not permitted |
| Outcomes | Reported in paper: live birth, ongoing pregnancy, clinical pregnancy, number of oocytes, total dose of FSH, duration of FSH, moderate and severe OHSS, had at least 1 transferable embryo, cycle cancellations for poor and hyper-response<br>Obtained from author correspondence*: multiple pregnancy<br>Not available: poor response, normal response, hyper-response |
| Notes | Trial registration: ISRCTN82461750 and EUCTR2006-001143-59-GB*, prospectively<br>Funding: Merck-Serono, unconditional research grant*<br>Conflict of interest: none declared<br>This study was also presented at the British Fertility Society 2009 - poster*<br>Email correspondence undertaken: yes, with corresponding author K Jayaprakasan ( Jayaprakasan 2017 [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| Bias | Authors' judgement | Support for judgement |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper states randomization was achieved with a "computer-generated pseudorandom code using random permuted blocks of varying size", and this was confirmed to be random by the clinical trials support unit staff (email correspondence) |
| Allocation concealment (selection bias) | Low risk | The allocation was concealed within the system until the point of randomization |
| Blinding of participants and personnel (performance bias)<br>All outcomes | High risk | The embryologists were blinded to the gonadotrophins dose that the participants had received; however, other study personnel such as doctors not blind, therefore high risk |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | High risk | The embryologists were blinded to the gonadotrophins dose that the participants had received; however, no other study personnel were blind, therefore high risk |

**Jayaprakasan 2010** (*Continued*)

| | | |
|---|---|---|
| Incomplete outcome data (attrition bias) All outcomes | Low risk | 2 women from each arm were excluded due to pregnancy and personal reasons; this is balanced across groups and low numbers |
| Selective reporting (reporting bias) | Low risk | Trial was registered prospectively and all outcomes are reported |
| Other bias | Low risk | - |

**Klinkert 2005**

| | |
|---|---|
| Methods | RCT,**direct dose comparison study**, 52 randomized<br>Study arms: 2<br>Setting: the Netherlands, 1 centre: University Medical Center Utrecht<br>Recruitment period: May 2001 to November 2002 |
| Participants | Population: anticipated **LOW** responders<br>Age: eligibility criteria < 47 years old. Median (10-90 percentile) G1: 40.4 years (36.6-44.5), G2: 42.2 years (33.7-44.6)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>**AMH**: no eligibility criteria, not recorded<br>**AFC**: eligibility criteria < 5. Median (10-90 percentile) G1: 3.0 (2.0-4.0), G2: 3.0 (0.7-4.0). AFC definition: follicles in both ovaries measuring 2-5 mm*<br>**bFSH (IU/L)**: no eligibility criteria. Median (10-90 percentile) G1: 9.3 IU/L (5.5-22.6), G2: 12.0 IU/L (5.8-20.8)<br>**Criteria relating to prior IVF/ICSI response**: no<br>Additional inclusion criteria: a regular spontaneous menstrual cycle of 25-35 days, the presence of both ovaries<br>Additional exclusion criteria: women with large ovarian cysts (> 30 mm) |
| Interventions | Group 1 dose/drug: **150 IU** (Gonal-F, Serono)<br>Group 2 dose/drug: **300 IU** (Gonal-F, Serono)<br>Protocol: long agonist<br>Dose titration: dose adjustment was permitted only in the 150 IU arm "the dose was doubled after 7 days of stimulation if the estradiol level was < 200 pmol/L or after 10 days if the estradiol level was < 500 pmol/L" |
| Outcomes | Reported in paper: ongoing pregnancy, clinical pregnancy, multiple pregnancy, cancellations for poor and hyper-response, poor response, total dose of FSH<br>Obtained from author correspondence: moderate and severe OHSS, had at least 1 transferable embryo, number of oocytes<br>Not available: live birth, normal response, hyper-response, duration of FSH |
| Notes | Trial registration: not registered*<br>Funding: unconditional research grant from Serono*<br>Conflict of interest:<br>Presented in part as an oral presentation at the 19th annual meeting of the ESHRE in Madrid, July 1, 2003 |

**Klinkert 2005** (*Continued*)

| | |
|---|---|
| | Email correspondence undertaken: yes, with Ellen Klinkert and Frank Broekmans (note Frank is an author on this review)(Broekmans 2017 [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| Bias | Authors' judgement | Support for judgement |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper contains no description; however, author correspondence reveals a "[c]omputer generated randomization list" was used |
| Allocation concealment (selection bias) | Low risk | Paper states participants "were randomized by opening a sealed envelope that contained information on the starting dose" and author correspondence elaborates, "[o]paque, and sealed envelopes, numbered with the study numbers" |
| Blinding of participants and personnel (performance bias)<br>All outcomes | High risk | Not blinded and dose-adjustment permitted in 1 arm. Paper states that 9 of the women who started with 150 IU, had to be increased to 300 IU due to an insufficient response, therefore high risk |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | High risk | No blinding employed therefore high risk |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | No attrition |
| Selective reporting (reporting bias) | High risk | Trial not registered, protocol not available, live birth not reported |
| Other bias | Low risk | - |

**Lan 2013**

| | |
|---|---|
| Methods | RCT, **ORT-algorithm study**, 348 randomized<br>Study arms: 2<br>Setting: Vietnam, 1 centre: An Sinh Hospital<br>Recruitment period: 1 October 2011 to 31 August 2012 |
| Participants | Age: eligibility criteria < 40 years. Mean (SD) G1: 32.3 years (4), G2: 33.1 years (4.1)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>**AMH (pmol/L)**: no eligibility criteria. Mean (SD) G1: 18.6 pmol/L (12.1), G2: 22.1 pmol/L (13.6) Assay: AMH gen II, Beckman Coulter |

| | |
|---|---|
| | **AFC**: no eligibility criteria: mean (SD) G1: 8.9 (4.8), G2: 11.2 (6.4) AFC definition: follicles in both ovaries measuring 2-9 mm*<br>**bFSH (IU/L)**: eligibility criteria < 12 IU/L. G1: 5.7 IU/L (2.5), G2: 5.8 IU/L (2.4)<br>**Criteria relating to prior IVF/ICSI response**: no<br>Additional inclusion criteria: BMI < 28 kg/$m^2$<br>Additional exclusion criteria: participation in another interventional clinical trial or concomitant use of either LH or human menopausal gonadotrophin/urinary FSH preparations in the study cycle |
| Interventions | Group 1: **AMH-tailored**: starting dosing of rFSH was 375 IU/day, 225 IU/day, or 150 IU/day in women having basal serum AMH concentrations of < 0.7, 0.7-2.1, 2.1 ng/mL, respectively<br>Group 2: **AFC-tailored**: starting dosing of rFSH was 375 IU/day, 225 IU/day, or 150 IU/day in participants having < 6, 6-15, > 15, respectively<br>Protocol: long agonist<br>Dose titration: adjustment was permitted in both arms after 5 days according to clinical judgement |
| Outcomes | Reported in paper: ongoing pregnancy, clinical pregnancy, moderate and severe OHSS, multiple pregnancy, number of oocytes, total dose of FSH, duration of FSH, cycle cancellations for poor and hyper-response, poor response, normal response, hyper-response<br>Obtained from author correspondence*: had at least 1 transferable embryo, cycle cancellations for poor and hyper-response<br>Not available: live birth data provided by authors but due to high loss to follow-up for this outcome we did not include it |
| Notes | Trial registration: NCT01783301, retrospectively<br>Funding: -<br>Conflict of interest: "The authors report no financial or commercial conflicts of interest"<br>Presented at ASPIRE 2014 meeting*<br>Email correspondence undertaken: yes, with corresponding author Vuong Thi Ngoc Lan (Lan 2017 [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| Bias | Authors' judgement | Support for judgement |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper states, "[p]atients were randomised by means of sealed envelopes generated by a computer randomization list" |
| Allocation concealment (selection bias) | Low risk | Paper states that sealed envelopes were used; however, there is no further description. Email correspondence confirms "the study nurse called the independent study coordinator, she opened the sealed envelope without knowing the allocation group inside" as third-party randomisation used, |

**Lan 2013** (*Continued*)

| | | |
|---|---|---|
| | | SNOSE envelopes not required, therefore low risk |
| Blinding of participants and personnel (performance bias)<br>All outcomes | High risk | Described as "open-label". The authors state via email that the "clinician making dose adjustment was not aware of AMH, AFC levels and the group allocation"*; however, it is unclear how this could be achieved in an open-label trial |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | High risk | Described as 'open-label' |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | Authors confirm no attrition* However author provided data on outcome of live birth, for which 7 pregnant women in the AMH arm and 11 in the AFC arm were lost to follow-up |
| Selective reporting (reporting bias) | High risk | Trial registered retrospectively after finishing recruitment, important outcomes such as live birth not reported |
| Other bias | Low risk | - |

**Lefebrve 2015**

| | |
|---|---|
| Methods | RCT, **direct dose comparison study**, 356 randomised<br>Study arms: 2<br>Setting: Canada, 1 centre, OVO clinic (a university-affiliated private IVF centre)<br>Recruitment period: October 2009 to September 2013 |
| Participants | Population: anticipated **LOW** responders<br>Age: eligibility criteria < 41 years. Median (IQR) G1: 37.9 years (35.0-39.5), G2: 37.8 years (34.6-39.5)<br>**Ovarian reserve test measures**<br>Participants were at risk of poor response defined as: < 5 oocytes, < 8 follicles, or cancellation in a previous IVF cycle with ≥ 300 IU/day, bFSH < 10 IU/L, AMH < 7.14 pmol/L, or AFC ≤ 8<br>**AMH (pmol/L)**: eligibility criteria < 7.14 pmol/L. Median (IQR) G1: 3.0 pmol/L (1.5-5.3), G2: 3.14 pmol/L (1.6-5.9) - converted from ng/mL Assay: ELISA test by Beckman Coulter*<br>**AFC**: eligibility criteria ≤8. Median (IQR) G1: 8 (6-11), G2: 9 (7-11). AFC defined as 2-9 mm both ovaries*<br>**bFSH (IU/L)**: eligibility criteria: < 20 IU/L. Median (IQR) G1: 8.7 IU/L (6.9-10.5), G2: 8.0 IU/L (6.5-10.0)<br>**Criteria relating to prior IVF/ICSI response**: yes, prior poor response an eligibility criteria |

**Lefebrve 2015** (*Continued*)

| | |
|---|---|
| | Additional inclusion criteria: aged < 41 years old, BMI < 35 kg/m$^2$, primary or secondary infertility and indicated for IVF/ICSI<br>Additional exclusion criteria: participation in other trial, women using or who have used investigational drugs in last 3 months, women with HIV, Hep B, Hep C |
| Interventions | Group 1 dose/drug: **450 IU** (225 IU Menopur and 225 IU Bravelle Ferring Pharmaceuticals)<br>Group 2 dose/drug: **600 IU** (300 IU Menopur and 300 IU Bravelle Ferring Pharmaceuticals)<br>Protocol: microdose agonist flare-up with 17b-E2 tablet priming<br>Dose titration: no dose-adjustment was permitted* |
| Outcomes | Reported in paper: clinical pregnancy, number of oocytes, total dose of FSH, duration of FSH<br>Obtained from author correspondence: live birth, multiple pregnancy, moderate and severe OHSS, cycle cancellations for low and hyper-response, had at least 1 transferable embryo<br>Not available: low-response, normal response, hyper-response |
| Notes | Trial registration: NCT00971152: prospectively registered<br>Funding: Ferring Pharmaceuticals Canada contributed to the study by supplying part of the medication. Ferring also assumed the costs of the statistician involved in designing the initial protocol and of the Ethics Committee fees<br>This study has been presented at the 2014 annual ESHRE meeting in Munich<br>Conflict of interest: industry funding<br>Email correspondence undertaken: yes with Jessica Lefebrve (Lefebvre 2017 [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| Bias | Authors' judgement | Support for judgement |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper states, "randomization was done by means of sequential study numbers (ratio 1:1)" and author correspondence confirmed randomisation "was generated by computer" |
| Allocation concealment (selection bias) | Low risk | Author correspondence confirmed "the envelopes were sealed white opaque envelopes. Each envelope had the randomization number and were opened in sequential order ... participant # 205 would open envelope # 205", therefore envelopes meeting SNOSE criteria |

**Lefebrve 2015** (*Continued*)

| | | |
|---|---|---|
| Blinding of participants and personnel (performance bias)<br>All outcomes | High risk | Described as "nonblinded" in paper |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | High risk | Described as "nonblinded" in paper |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | In each arm 5 participants did not commence IVF treatment: small number and reasons balanced |
| Selective reporting (reporting bias) | Low risk | Live birth, multiple pregnancy and OHSS included as outcomes on trial registration but not reported in paper. However, author provided these outcome data by email |
| Other bias | Low risk | - |

**Magnusson 2017**

| | |
|---|---|
| Methods | RCT, **ORT-algorithm study**, 308 randomised<br>Study arms: 2<br>Setting: Sweden, 1 centre: Sahlgrenska University Hospital<br>Recruitment period: January 2013 to May 2016 |
| Participants | Age: eligibility criteria: > 18 and < 40 years. Mean (SD) G1: 32.3 years (4.0), G2: 32.3 years (3.8)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>**AMH (pmol/L)**: no eligibility criteria. For AMH group only: mean (SD) G1: 28.8 pmol/L (25.2) G2: not provided - converted from ng/mL. Assay: modified Beckman Coulter Hen II assay<br>**AFC**: no eligibility criteria: mean (SD) G1: 21.6 (12.0), G2: 21.3 (11.3). AFC definition: follicles in both ovaries measuring 2-10 mm<br>**bFSH**: no eligibility criteria, not recorded*<br>**Criteria relating to prior IVF/ICSI response**: no<br>Additional inclusion criteria: BMI above 18.0 kg/$m^2$ and below 35.0 kg/$m^2$, having their first standard IVF planned and AMH not previously measured<br>Additional exclusion criteria: male factor infertility where ICSI was planned, cycles planned for oocyte donation or PGD |
| Interventions | Group 1: **starting dose determined by an algorithm including AFC,age and BMI** (no AMH)(Gonal-F, Merck)<br>Group 2: **starting dose determined by an algorithm including AMH,AFC,age and BMI** (Gonal-F, Merck)<br>Protocol: long agonist<br>Dose titration: dose adjustment was allowed at the earliest on day 7 and only in predefined steps and if E2 on stimulation day 6 was either < 350 pmol/L or > 1500 pmol/L. |

**Magnusson 2017** (*Continued*)

| | |
|---|---|
| | Investigators performing dose-adjustments were blind to allocation |
| Outcomes | Reported in paper: live birth, ongoing pregnancy, number of oocytes, total dose of FSH, moderate and severe OHSS, poor response, normal response, hyper-response<br>Obtained from author correspondence*: multiple pregnancy, clinical pregnancy, cycle cancellations for poor and hyper-response, had at least 1 transferable embryo, severe OHSS, duration of FSH<br>This study included fresh transfer, and for women with a freeze-all the first frozen transfer could be counted |
| Notes | Trial registration: NCT02013973, registered during recruitment*<br>Funding: Ferring Pharmaceuticals (unrestricted grant), Sahlgrenska University Hospital, Hjalmar Svensson Research Foundation<br>Conflict of interest: none declared<br>No conference presentation*<br>Email correspondence undertaken: yes with corresponding author Åsa Magnusson ( Magnusson 2017 [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| Bias | Authors' judgement | Support for judgement |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper states, "[r]andomization was performed with a computerized randomization program ... in the proportions of 1:1" |
| Allocation concealment (selection bias) | Low risk | Paper states "... randomization program with concealed allocation of patients" |
| Blinding of participants and personnel (performance bias)<br>All outcomes | Low risk | Paper states that "[t]he study was blinded to patients, physicians managing patients during IVF treatment and the statistician. For practical reasons, the two physicians performing the AFC estimations were unblinded to the starting dose of recombinant FSH (rFSH)", therefore no potential for performance bias |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | Low risk | Paper states that "[t]he study was blinded to patients, physicians managing patients during IVF treatment and the statistician. For practical reasons, the two physicians performing the AFC estimations [prior to randomization] were unblinded to the starting dose of recombinant FSH (rFSH)", therefore no potential for detection bias |

**Magnusson 2017** (*Continued*)

| | | |
|---|---|---|
| Incomplete outcome data (attrition bias) All outcomes | Low risk | 1 participant excluded postrandomisation |
| Selective reporting (reporting bias) | Low risk | Study was registered during recruitment, but prior to the end of the trial. All outcomes reported including live birth |
| Other bias | Low risk | - |

**Nyboe Andersen 2017**

| | |
|---|---|
| Methods | RCT (non-inferiority design), **ORT-algorithm study**, 1329 randomised<br>Study arms: 2<br>Setting: 7 investigational sites in 11 countries (Belgium, Brazil, Canada, Czech Republic, Denmark, France, Italy, Poland, Russia, Spain, and UK)<br>Recruitment period: 8 October 2013 to 11 May 2015 |
| Participants | Age: eligibility criteria: 18-40 years. Mean (SD) G1: 33.4 years (3.9), G2: 33.2 years (3.9)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>**AMH (pmol/L)**: no eligibility criteria. Median (IQR): G1: 16.3 pmol/L (9.0-24.8) , G2: 16.0 pmol/L (9.1-25.5). Assay: automated Elecsys AMH immunoassay (Roche Diagnostics International)<br>**AFC**: no eligibility criteria. Mean (SD) G1: 14.7 (6.9), G2: 14.4 (6.8). AFC definition: follicles in both ovaries measuring 2-10 mm<br>**bFSH (IU/L)**: eligibility criteria: 1-15 IU/L. Median (IQR): G1: 7.5 IU/L(6.2-9.2), G2: 7.7 IU/L (6.5-9.4)<br>**Criteria relating to prior IVF/ICSI response**: no<br>Additional inclusion criteria: undergoing their first IVF/ICSI cycle and diagnosed with unexplained infertility, tubal infertility, endometriosis stage I/II, or with partners diagnosed with male factor infertility, BMI 17.5-32.0 kg/m$^2$, regular menstrual cycles of 24-35 days, presence of both ovaries<br>Additional exclusion criteria: endometriosis stage III-IV, history of recurrent miscarriage, and use of hormonal preparations (except for thyroid medication) |
| Interventions | Group 1: standard fixed dose **150 IU** (Gonal-f, EMD Serono)<br>Group 2: individualised dose **based on AMH and body weight**:<br>AMH : dose in µg or µg/kg:<br>• < 15: 12 µg<br>• 15-16: 0.19 µg/kg<br>• 17: 0.18 µg/kg<br>• 18: 0.17 µg/kg<br>• 19-20: 0.16 µg/kg<br>• 21-22: 0.15 µg/kg<br>• 23-24: 0.14 µg/kg<br>• 25-27: 0.13 µg/kg<br>• 28-32: 0.12 µg/kg<br>• 33-39: 0.11 µg/kg |

| | |
|---|---|
| | • ≥ 40: 0.10 µg/kg (Follitropin delta, Ferring Pharmaceuticals)<br>Protocol: antagonist<br>Dose titration: women in the 150 IU arm were administered 150 IU for the first 5 days, thereafter the dose could be adjusted up or down according to follicular response, with 450 IU as the maximum daily dose allowed. Dose-adjustment was not permitted in the individualised dosing arm |
| Outcomes | Reported in paper: live birth, ongoing pregnancy, clinical pregnancy, number of oocytes, total dose of FSH, duration of FSH, multiple pregnancy, moderate and severe OHSS, had at least 1 transferable embryo, normal response to stimulation<br>Obtained from author correspondence*: cycle cancellations for poor and hyper-response<br>Not available: severe OHSS, poor response, hyper-response (this data was available only in participant subgroups, and authors declined to provide the non-stratified data) |
| Notes | Trial registration: NCT01956110, prospectively<br>Funding: Ferring Pharmaceuticals<br>Conflict of interest: many investigators have received grants or are employed by Industry<br>Presented in part at the 32nd Annual Meeting of the European Society of Human Reproduction and Embryology, July 3-6, 2016, Helsinki, Finland<br>Email correspondence undertaken: yes with Ferring staff; however, not all requested information was provided (Helmgaard 2017b [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| Bias | Authors' judgement | Support for judgement |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper states "Women were randomly assigned in a 1:1 ratio via a central computer-generated randomization sequence, prepared by an independent statistician. Randomization was stratified by age (< 35, 35-37, and 38-40 years) and performed in blocks of four within trial sites." |
| Allocation concealment (selection bias) | Low risk | No description in the paper; however, correspondence with Ferring reveals that the randomisation allocation was concealed within the computer programme until the time of randomisation. The unblinded study nurse managed the randomisation as the other investigators were unaware of the trial allocations* |
| Blinding of participants and personnel (performance bias)<br>All outcomes | Unclear risk | Paper states, "[a]ll investigators, embryologists, and central laboratory personnel were blinded to treatment allocation" and "[i]nvestigators evaluated the need for dose adjustments in a treatment blinded manner |

| | | |
|---|---|---|
| | | on the basis of follicular development, and requests for dose increases or decreases were implemented as applicable by an unblinded study nurse" This description suggests the investigators making decisions about dose adjustment did not speak with the participants, therefore low risk; however, there are other means by which performance bias may operate than dose-adjustment. As the participants were not blind they may have disclosed their dose allocation to the staff, and it is unclear whether there were any safeguards to prevent this |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | Unclear risk | Paper states, "[a]ll investigators, embryologists, and central laboratory personnel were blinded to treatment allocation." However, the participants were not blind and may have disclosed their dose allocation to the staff, and it is unclear whether there were any safeguards to prevent this |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | 3 women were excluded from the analysis as were not exposed to the study drug; small number in a large trial therefore low risk |
| Selective reporting (reporting bias) | High risk | The authors present the rate of poor response in the low AMH group and the rate of hyper-response in the high AMH group but do not present the rate of poor response in the high AMH group or the rate of hyper-response in the low AMH group. This also appears to be a departure from the analysis plans listed on the trial registration website. Although the authors state they had always planned to analyse the data this way, the data leave open the possibility that the individualised regimen may increase excessive responses in low AMH participants |
| Other bias | High risk | Interim analysis with no apparent p-value correction performed. May produce biased data for the purposes of a systematic review |

**Olivennes 2015**

| | |
|---|---|
| Methods | RCT, **ORT-algorithm study**, 200 randomised<br>Study arms: 2<br>Setting: 22 centres across 9 European countries and 1 centre in Chile<br>Recruitment period: August 2008 to January 2010 |
| Participants | Age: eligibility criteria: 18-34 years. Mean (SD) G1: 30.6 years (2.6), G2: 30.0 years (2.9)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>**AMH**: no eligibility criteria: mean (SD) G1: 17.1 pmol/L (10.7) G2: 19.3 (13.6) pmol/L (converted from ng/L). Assay: Quest Nichols (Specialty)/Q2 (ORL 22)*<br>**AFC**: no eligibility criteria: mean (SD) G1: 16.0 (7.2), G2: 17.6 (7.2). AFC definition: follicles in both ovaries measuring 2-11 mm<br>**bFSH (IU/L)**: eligibility criteria < 12 IU/L. Mean (SD) G1: 6.8 IU/L (1.84), G2: 6.8 IU/L (1.53)<br>**Criteria relating to prior IVF/ICSI response**: yes, previous poor response to ovarian stimulation (defined as 5 or fewer mature follicles or 3 or fewer oocytes collected) in 2 or more assisted reproductive technique cycles; previous hyper-response to ovarian stimulation (defined as 25 or more oocytes retrieved) in 2 or more assisted reproduction technique cycles; or previous severe OHSS<br>Additional inclusion criteria: a regular spontaneous menstrual cycle of 21-35 days; BMI less than 30 kg/m$^2$; and a male partner with semen analysis (within the past 6 months) considered adequate for regular IVF/ICSI donor sperm was required if the partner's semen analysis was considered inadequate<br>Additional exclusion criteria: 3 or more spontaneous abortions; PCOS, endometriosis or uterine fibroids that require treatment; or any other medical condition that may have affected the absorption,distribution,metabolism or excretion of follitropin alfa |
| Interventions | Group 1: **150 IU** (Gonal-F, Merck Serono)<br>Group 2: **CONSORT algorithm**: determined by women's age, height, weight, serum FSH level and AFC (CONSORT calculator). Participants were assigned to 1 of the following rFSH dose groups: 112.5, 150, 187.5, 225, 300, or 450 IU per day (Gonal-F, Merck Serono)<br>Protocol: long agonist<br>Dose titration: in G1 the dose was maintained for the first 5 days. Then the dose could be increased or decreased depending on the woman's ovarian response. In G2 the allocated dose was maintained throughout the treatment cycle unless a patient was considered by the investigator to be at risk of OHSS. In such cases, the dose was decreased |
| Outcomes | Reported in paper: live birth, ongoing pregnancy, clinical pregnancy, number of oocytes, total dose of FSH, duration of FSH, multiple pregnancy, had at least 1 transferable embryo, cycle cancellations for poor and hyper-response<br>Obtained from author correspondence*: moderate and severe OHSS<br>Not available: poor response, normal response, hyper-response |
| Notes | Trial registration: NCT00829244, registered approximately 5 months after recruitment commenced<br>Funding: Merck Serono SA*<br>Conflict of interest:<br>No relevant conference presentation* |

**Olivennes 2015** (*Continued*)

| | |
|---|---|
| | Email correspondence undertaken: yes, with Merck Serono (Alam 2017 [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| Bias | Authors' judgement | Support for judgement |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper states "Patients were randomised (1:1 ratio; stratified by centre) to receive rFSH according to either CONSORT or standard dosing. An electronic case report form system was used to allocate patients based on a list prepared for each centre to ensure that the randomisation across the two groups was balanced in a 1:1 ratio" |
| Allocation concealment (selection bias) | Low risk | Not stated in paper but author correspondence confirmed randomisation was automatically generated by the computer-system upon request, so allocation was concealed* |
| Blinding of participants and personnel (performance bias)<br>All outcomes | High risk | Stated as open label "Patients and investigators were aware of the allocated treatment group and rFSH dose" |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | High risk | Stated as open label "Patients and investigators were aware of the allocated treatment group and rFSH dose |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | 1 participant who committed a protocol violation was excluded |
| Selective reporting (reporting bias) | Low risk | Trial registered during recruitment but prior to finishing the trial, all registered outcomes reported including live birth |
| Other bias | Low risk | - |

**Oudshoorn 2017**

| | |
|---|---|
| Methods | RCT, **direct dose comparison study**, 521 randomised<br>Study arms: 2<br>Setting: 25 academic and nonacademic centres in the Netherlands*<br>Recruitment period: May 2011 and May 2014<br>RCT was embedded in a Dutch cohort study (with Van Tilborg 2017) |

**Oudshoorn 2017** (*Continued*)

| | |
|---|---|
| Participants | Population: anticipated **HIGH** responders<br>Age: < 44 Mean (SD) G1: 31.6 years (4.5), G2: 32.0 years (4.3)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>**AMH**: no eligibility criteria: median (IQR) G1: 23.8 pmol/L (1.98), G2: 3.00 pmol/L (1.89)(converted from ng/ml)<br>**AFC**: eligibility criteria: AFC > 15. Median (IQR) G1: 20.0 (7.0), G2: 21.0 (8.0). AFC definition: follicles in both ovaries measuring 2-10 mm<br>**bFSH**: not reported*<br>**Criteria relating to prior IVF/ICSI response**: no, women undergoing first IVF/ICSI cycle<br>Additional inclusion criteria: undergoing first IVF/ICSI cycle with a regular indication for IVF/ICSI, regular cycle (average cycle length of 25-35 days)<br>Additional exclusion criteria: PCOS (Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group) |
| Interventions | Group 1: **150 IU** (variable manufacturers)<br>Group 2: **100 IU** (variable manufacturers)<br>Protocol: agonist or antagonist<br>Dose titration: dose adjustments during stimulation were not allowed. Between treatment cycle dose adjustments were allowed in both study arms following strict, pre-determined criteria. In the reduced dose group the FSH dose could be adjusted with a step of 25 IU in case of a poor or hyper-response. Poor response was defined as the cancellation of a stimulation cycle if < 2 follicles > 12 mm in diameter or < 3 follicles > 17 mm were observed on TVS or if < 5 oocytes were retrieved. Hyper-response was defined as cancellation of a stimulation cycle because > 20 follicles of > 12 mm in diameter were growing and estradiol levels exceeded 11.700 pmol/L (3187.08 ng/L), if > 30 follicles of > 12 mm were growing or if > 15 oocytes were retrieved. For the standard dose group a dose adjustment between cycles was allowed with a maximum of 50 IU FSH, following the criteria mentioned above |
| Outcomes | Reported in paper: live birth, ongoing pregnancy, poor response, hyper-response (normal response calculated) cycle cancellations for poor and hyper-response, number of oocytes<br>Obtained from author correspondence*: clinical pregnancy (for first IVF cycle), multiple pregnancy (for first IVF cycle), had at least 1 transferable embryo, total dose of FSH, duration of FSH, moderate and severe OHSS<br>Not available: none |
| Notes | Trial registration: NTR2657 (Dutch register)<br>Funding: ZonMw, the Dutch Organization for Health Research and Development<br>Conflict of interest: trial collaborators are authors of this review. HT received an unrestricted research grant from Merck Serono (the Netherlands). The Department of Obstetrics and gynaecology, University Medical Centre Groningen receives an unrestricted research grant from Ferring pharmaceutics BV (the Netherlands). FB receives monetary compensation as a member of the external advisory board for Ferring pharmaceutics BV and Merck Serono for consultancy work for Gedeon Richter (Belgium) and Roche Diagnostics (Switzerland) and for a research cooperation with Ansh Labs (USA). BM reports consultancy for OvsEva, Merck and Guerbet<br>The study was presented at ESHRE 2016 (O-035, O-036, O-037)<br>Email correspondence undertaken: yes, with Helen Torrance (review author)(Torrance |

**Oudshoorn 2017** (*Continued*)

| | |
|---|---|
| | 2017 [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| Bias | Authors' judgement | Support for judgement |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Stated as randomisation "using a web-based randomization program" |
| Allocation concealment (selection bias) | Low risk | Not stated in paper but authors confirmed via email that the trial allocation was only revealed after entering information regarding the participants' eligibility criteria on the web-based system and then clicking 'randomise' |
| Blinding of participants and personnel (performance bias)<br>All outcomes | High risk | Stated as "open-label" |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | High risk | Stated as "open-label" |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | Only 2 participants with missing data for our primary outcomes. Loss to follow-up was also minimal in both groups over the 18-month period (16/255 (6.3%) in the reduced dose group vs 18/266 200 (6.8%) in the standard dose group) |
| Selective reporting (reporting bias) | Low risk | Study protocol published and all outcomes reported |
| Other bias | Low risk | - |

**Out 2004**

| | |
|---|---|
| Methods | RCT, **direct dose comparison study**, 164 randomised<br>Study arms: 2<br>Setting: UK (6 centres)<br>Recruitment period: June 2000 to December 2001 |
| Participants | Population: anticipated **NORMAL** responders<br>Age: eligibility criteria: 18-39 years. Mean (SD) G1: 32.7 years (3.6), G2: 32.2 years (3.5)<br>**Ovarian reserve test**: e**ligibility criteria and demographics**<br>**AMH**: no eligibility criteria, not recorded* |

| | |
|---|---|
| | **AFC**: no eligibility criteria, not recorded*<br>**bFSH (IU/L)**: eligibility criteria: elevated early follicular phase (menstrual cycle day 2 ± 7) circulating FSH and/or LH concentrations according to cutoff levels used in the local laboratory. Mean (SD) G1: 6.3 IU/L (1.8), G2: 6.1 IU/L (1.6)<br>**Criteria relating to prior IVF/ICSI response**: no<br>Additional inclusion criteria: normal regular menstrual cycles with a range of 24-35 days; BMI 18-29 kg/m$^2$; and body weight 50-90 kg<br>Additional exclusion criteria: history of or current endocrine abnormality such as PCOS or evidence of ovarian dysfunction; any clinically significant abnormal laboratory value; any ovarian and/or abdominal abnormality that would interfere with adequate ultrasound investigation of at least 1 ovary; only 1 ovary; contra-indications for the use of gonadotropins; use of hormonal preparations within 1 month prior to the date of signing consent; alcohol or drug abuse, or history thereof, within the 12 months preceding signing informed consent; or administration of investigational drugs within 3 months prior to screening |
| Interventions | Group 1 dose/drug: **150 IU** (follitropin ß, Organon)<br>Group 2 dose/drug:**200 IU** (follitropin ß, Organon)<br>Protocol: antagonist<br>Dose titration: "Dose of rFSH could be adjusted downwards to 100 IU daily based on the clinical judgment of the investigator. For this purpose separate vials containing 100 IU were made available" - as the trial was blinded these adjustments were made blind to the participants' allocation |
| Outcomes | Reported in paper: clinical pregnancy, moderate and severe OHSS, number of oocytes retrieved, total dose of FSH, duration of FSH, cycle cancellations for poor and hyper-response, had at least 1 transferable embryo<br>Obtained from author correspondence: none<br>Not available: live birth, ongoing pregnancy, multiple pregnancy, poor response, normal response, hyper-response |
| Notes | Trial registration: none stated<br>Funding: Organon, who assisted with study monitoring<br>Conflict of interest: none stated<br>Email correspondence undertaken: yes with Geoffrey Trew, co-author (Trew 2017 [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| Bias | Authors' judgement | Support for judgement |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper states participants were "randomised by receiving a subject number from a randomization list corresponding with patient boxes in which the medication was kept" "The randomization was done in blocks of four and was computer-generated using random numbers." |

**Out 2004** (*Continued*)

| | | |
|---|---|---|
| Allocation concealment (selection bias) | Low risk | Study is double-blind therefore allocation concealment ensured |
| Blinding of participants and personnel (performance bias)<br>All outcomes | Low risk | Study described as double-blind and "[t]he 150 and 200 IU rFSH vials were indistinguishable", therefore all parties blind |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | Low risk | Study described as double-blind and "[t]he 150 and 200 IU rFSH vials were indistinguishable", therefore all parties blind |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | Appears to be no attrition |
| Selective reporting (reporting bias) | High risk | Trial not registered, no protocol available, and not reporting important outcomes such as live birth |
| Other bias | Low risk | - |

**Popovic-Todorovic 2003**

| | |
|---|---|
| Methods | RCT, **ORT-algorithm study**, 262 randomised<br>Study arms: 2<br>Setting: Denmark, 2 centres: Rigshospitalet and Hvidovre hospitals<br>Recruitment period: January 2002 and January 2003 |
| Participants | Age: eligibility criteria: < 39 years. Mean (SD) G1: 31.9 years (3.9), G2: 32.7 years (3.7)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>**AMH**: no eligibility criteria, not reported<br>**AFC**: no eligibility criteria, Mean (SD): 18.84 (7.6)*. AFC definition: follicles in both ovaries measuring 5-10 mm<br>**bFSH (IU/L)**: eligibility criteria 12.5 IU/L. Mean (SD): 6.95 IU/L (1.75)*<br>**Criteria relating to prior IVF/ICSI response**: no<br>Additional inclusion criteria: first IVF/ICSI treatment cycle; presence of both ovaries; regular spontaneous menstrual cycle (21-35 days); no evidence of endocrine disorders<br>Additional exclusion criteria: presence of ovarian cysts and inaccessible ovaries |
| Interventions | Group 1: **150 IU** (Puregon, Organon)<br>Group 2: **100 to 250 IU based on a normogram** consisting of AFC, total Doppler score on days 2-5, total ovarian volume on days 2-5, age, and smoking status (Puregon, Organon)<br>Protocol: long agonist<br>Dose titration: "Dose adjustments were allowed after day 8 of stimulation. The dose was increased if the leading follicles were < 10±11 mm and in case of asynchrony (i.e. more than 4 mm difference between the leading follicle and the next pool). The rFSH dose was reduced if a risk of developing an excessive number of follicles (> 20) was acknowledged" |

**Popovic-Todorovic 2003** (*Continued*)

| | |
|---|---|
| Outcomes | Reported in paper: clinical pregnancy, ongoing pregnancy, number of oocytes, total dose of FSH, duration of FSH, cycle cancellations for poor and hyper-response, poor response, normal response, hyper-response, had at least 1 transferable embryo<br>Obtained from author correspondence*: multiple pregnancy, moderate and severe OHSS<br>Not available: live birth |
| Notes | Trial registration: confirmed not registered*<br>Funding: Organon provided the PhD grant for the first author*<br>Conflict of interest: nothing stated<br>The study was presented at ESHRE 2003 in Madrid*<br>Email correspondence undertaken: yes, with corresponding author Biljana Popovic-Todorovic (Popovic-Todorovic 2017 [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| **Bias** | **Authors' judgement** | **Support for judgement** |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper states "patients were randomised via computer-generated lists using "clusters of 10" |
| Allocation concealment (selection bias) | Low risk | Paper states, "[t]he randomization system was open, but handled independently of the clinicians treating the patients" which is unclear; however, the authors confirmed that third party randomisation was used and the clinician requesting the next randomisation code was not aware of the next allocation in advance* |
| Blinding of participants and personnel (performance bias)<br>All outcomes | High risk | No blinding described; authors declare potential for bias in absence of blinding for dose-adjustment decisions |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | High risk | No blinding described; authors declare potential for bias in absence of blinding for dose-adjustment decisions |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | Authors confirm no study attrition occurred* |
| Selective reporting (reporting bias) | High risk | Trial not registered, no protocol available. Live birth and a number of review outcomes not reported, therefore potential for selection bias |
| Other bias | Low risk | - |

**Tan 2005**

| | |
|---|---|
| Methods | RCT, **direct dose comparison study**, 192 randomised<br>Study arms: 2<br>Setting: Canada, 6 centres<br>Recruitment period: unknown |
| Participants | Population: anticipated **NORMAL** responders<br>Age: eligibility criteria: 18-39 years old. Mean (SD) G1: 33.3 years (3.1), G2: 33.4 years (3.3)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>Eligible if bFSH level was 'normal' (no definition provided)<br>**AMH (pmol/L)**: not reported<br>**AFC**: not reported<br>**bFSH (IU/L)**: eligibility criteria: 'normal'. Demographics not reported.<br>**Criteria relating to prior IVF/ICSI response**: yes, excluded if previous ovarian stimulation cycles in which fewer than 3 oocytes were retrieved<br>Additional inclusion criteria: cause of infertility potentially treatable by IVF or ICSI; normal ovulatory cycles with a mean cycle length of between 24 and 35 days; good physical and mental health; BMI 18-29 kg/$m^2$<br>Additional exclusion criteria: infertility caused by endocrine abnormalities such as hyperprolactinaemia, PCOS, absence of ovarian function; chronic cardiovascular, hepatic, renal or pulmonary disease; either current or previous (within 12 months) alcohol or drug abuse; administration of any investigational drugs within 3 months prior to screening |
| Interventions | Group 1 dose/drug: **100 IU** (Puregon; Organon)<br>Group 2 dose/drug: **200 IU** (Puregon; Organon)<br>Protocol: long agonist<br>Dose titration: dose adjustment was permitted after 4 days of FSH administration |
| Outcomes | Reported in paper: ongoing pregnancy, number of oocytes, total dose of FSH, moderate and severe OHSS, had at least 1 transferable embryo, cycle cancellations for poor and hyper-response<br>Obtained from author correspondence: none<br>Not available: live birth, poor response, normal response, hyper-response, multiple pregnancy, duration of FSH (means available but not SD), clinical pregnancy |
| Notes | Trial registration: none<br>Funding: Organon Canada provided the medication, statistical analysis, and support for the trial<br>Conflict of interest: none declared<br>Email correspondence undertaken: attempted but no reply received |

***Risk of bias***

| Bias | Authors' judgement | Support for judgement |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper states, "[e]ligible subjects were randomized by receiving a subject number from a randomization list corresponding |

| | | |
|---|---|---|
| | | with patient boxes in which the medication was kept. The randomization was carried out in blocks of four according to random numbers generated by the computer" therefore computer randomisation was used to number the boxes |
| Allocation concealment (selection bias) | Low risk | Allocation concealment is ensured by the double-blind nature of the trial: "The ampules used in the study were individually numbered for each subject. After allocation of subject code number, each subject used medications with the same code number throughout the study" |
| Blinding of participants and personnel (performance bias)<br>All outcomes | Low risk | The trial was double-blind: "Puregon (recombinant FSH) was supplied in 50 and 100 IU ampoules as lyophilized spheres. The two different dosage ampules appeared identical"<br>The participants and investigators were blind: "The investigator had no knowledge regarding the treatment assigned therefore the study was performed as a double-blind trial"<br>However the paper states that "[a]fter day 4 of stimulation, the r-FSH dose was adjusted if deemed necessary but the initial Puregon dose received was not revealed. The treatment cycle was no longer, from that point forward, assessor or patient blind" This implies that after day 4 there was no blinding of participants and personnel, and therefore there may be performance bias in the decision to cancel cycles, etc |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | High risk | As above, although the trial is described as double-blind it appears that unblinding occurred as only as day 4 of stimulation, and therefore assessment of OHSS would not be blind |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | Paper states that "185 patients completed the study as per study termination record (7 and 3 in the 100 IU and 200 IU group, respectively)"; however, it appears these were women cancelled during the cycle rather than withdrawn (either way the numbers remain low) |

**Tan 2005** (*Continued*)

| | | |
|---|---|---|
| Selective reporting (reporting bias) | High risk | Study not registered, no protocol available, live birth not reported |
| Other bias | Low risk | The recruitment target was 200 but only 192 women were recruited due to time-constraints, therefore power may be compromised; however, only marginally, therefore this was given a low risk rating |

**Tasker 2010**

| | |
|---|---|
| Methods | RCT, **ORT-algorithm study**, 286 randomised as per abstract (215 available for data analysis)<br>Study arms: 2<br>Setting: unclear<br>Recruitment period: not stated |
| Participants | Age: eligibility criteria: not stated<br>**Ovarian reserve test: eligibility criteria and demographics**<br>No description (abstract only)<br>**AMH**: eligibility criteria: unclear, not reported (however used in algorithm)<br>**AFC**: eligibility criteria: unclear, not reported (however used in algorithm)<br>**bFSH**: eligibility criteria: unclear, not reported (however used in algorithm)<br>**Criteria relating to prior IVF/ICSI response**: unclear<br>Additional inclusion criteria: first IVF/ICSI cycle<br>Additional exclusion criteria: unclear |
| Interventions | Group 1 dose/drug: dose calculated depending on women's age, early follicular FSH, E2 and the presence of polycystic ovaries<br>Group 2 dose/drug: dose calculated based on the AMH level and AFC, in addition to standard markers by using tables that were drawn up based on previous publications<br>Protocol: long agonist<br>Dose titration: unclear |
| Outcomes | Reported in abstract: poor response, normal response, hyper-response, ongoing pregnancy, total dose of FSH, duration of FSH<br>Obtained from author correspondence/individual participant data*: live birth, clinical pregnancy, number of oocytes<br>Not available: multiple pregnancy, moderate and severe OHSS, had at least 1 transferable embryo, cycle cancellations for poor and hyper-response,<br>All data was re-calculated using individual participant data, so may differ from that reported in abstract |
| Notes | Abstract only, and individual participant data provided by authors<br>Trial registration: unclear<br>Funding: unclear<br>Conflict of interest: nothing stated<br>Email correspondence undertaken: yes, and authors provided individual participant data; |

**Tasker 2010** (*Continued*)

| | |
|---|---|
| | however, were unable to assist with further questions (Hamoda 2017 [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| Bias | Authors' judgement | Support for judgement |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Abstract states, "[r]andomisation was achieved using a computer-generated list of numbers" |
| Allocation concealment (selection bias) | Low risk | Abstract states the allocations were "printed and placed into sequentially numbered opaque sealed envelopes (SNOSE)" |
| Blinding of participants and personnel (performance bias)<br>All outcomes | High risk | Blinding not stated, assumed nonblinded |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | Low risk | Blinding not stated, assumed nonblinded; however, no subjective outcomes reported - therefore this domain is not applicable |
| Incomplete outcome data (attrition bias)<br>All outcomes | High risk | 22 participants appear to have been withdrawn from the study based on the individual participant data |
| Selective reporting (reporting bias) | High risk | Trial does not appear to be registered, and unable to extract important data such as OHSS |
| Other bias | Unclear risk | Study available as abstract only, therefore much of the study methodology, etc. remains unclear. Individual participant data provided indicates lots of missing data, and a number of participants do not appear to have completed the trial at the time of final data collection; however, this is not considered an interim analysis, as there is no ongoing data collection |

**Van Tilborg 2017**

| | |
|---|---|
| Methods | RCT, **direct dose comparison study**, 511 (234+277) randomised<br>Study arms: 4 (treated as 2 separate trials for this review)<br>Setting: 25 academic and nonacademic centres in the Netherlands*<br>Recruitment period: May 2011 and May 2014<br>RCT was embedded in a Dutch cohort study (with Oudshoorn 2017) |
| Participants | This study stratified women based on AFC (0-7 and 8-10), and is essentially treated as 2 studies in this review:<br>Population: anticipated **LOW** responders<br>Age: < 44 years. Mean (SD) G1: 36.3 years (4.2), G2: 36.8 years (3.0)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>**AMH**: no eligibility criteria: median (IQR) G1: 6.0 pmol/L (6.14), G2: 4.6 pmol/L (5.2)<br>**AFC**: eligibility criteria: AFC 0-7. Median (IQR) G1: 6.0 (2.0), G2: 6.0 (3.0) AFC definition: follicles in both ovaries measuring 2-10 mm<br>**bFSH**: not reported*<br>**Criteria relating to prior IVF/ICSI response**: no, women undergoing first IVF/ICSI cycle<br>Additional inclusion criteria: undergoing first IVF/ICSI cycle with a regular indication for IVF/ICSI, regular cycle (average cycle length of 25-35 days)<br>Additional exclusion criteria: PCOS (Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group)<br>Population: anticipated **LOW** responders<br>Age: < 44 years, Mean (SD), G1: 35.1 years (4.2) G2: 35.0 years (4.7)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>**AMH (ng/mL)**: no eligibility criteria: median (IQR) G1: 1.17 ng/mL (0.4) G2: 1.13 ng/mL (1.05)<br>**AFC**: eligibility criteria: AFC 8-10. Median (IQR) G1: 9.0 (2.0) G2: 9.0 (2.0) AFC definition: follicles in both ovaries measuring 2-10 mm<br>**bFSH**: not reported*<br>**Criteria relating to prior IVF/ICSI response**: no, women undergoing first IVF/ICSI cycle<br>Additional inclusion criteria: undergoing first IVF/ICSI cycle with a regular indication for IVF/ICSI, regular cycle (average cycle length of 25-35 days)<br>Additional exclusion criteria: PCOS (Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group) |
| Interventions | AFC 0-7 trial:<br>• Group 1: **450 IU** (variable manufacturers)<br>• Group 2: **150 IU** (variable manufacturers)<br>AFC 8-10 trial:<br>• Group 1: **225 IU** (variable manufacturers)<br>• Group 2: **150 IU** (variable manufacturers)<br>Protocol: agonist or antagonist<br>Dose titration: dose adjustments during stimulation were not allowed. Between treatment cycle dose adjustments were allowed in the 150 IU arm only following strict, predetermined criteria: a maximum dose adjustment of 50 IU/day was allowed between cycles if women had a poor response (i.e., cycle cancellation due to insufficient growth: 167 < 2 follicles > 12 mm or < 3 follicles ≥ 17 mm; or < 5 oocytes at retrieval) or hyper- |

| | |
|---|---|
| | response (i.e. cycle 168 cancellation due to excessive response: > 20 follicles > 12 mm and estradiol levels exceeding 11.700 169 pmol/L (3187.08 ng/L) or > 30 follicles > 12 mm; or > 15 oocytes at retrieval) |
| Outcomes | Reported in paper: live birth, ongoing pregnancy, poor response, hyper-response (normal response calculated), cycle cancellations for poor and hyper-response, number of oocytes<br>Obtained from author correspondence*: clinical pregnancy (for first IVF cycle), multiple pregnancy (for first IVF cycle), had at least 1 transferable embryo, total dose of FSH, duration of FSH, moderate and severe OHSS<br>Not available: none<br>Data in paper was available in many instances only as merged 225/450 IU vs 150 IU, authors provided data split into the 2 sub-studies |
| Notes | Trial registration: NTR2657 (Dutch register)<br>Funding: ZonMw, the Dutch Organization for Health Research and Development<br>Conflict of interest: trial collaborators are authors of this review. HT received an unrestricted research grant from Merck Serono (the Netherlands). The Department of Obstetrics and Gynaecology, University Medical Centre Groningen receives an unrestricted research grant from Ferring Pharmaceutics BV (the Netherlands). FB receives monetary compensation as a member of the external advisory board for Ferring Pharmaceutics BV and Merck Serono for consultancy work for Gedeon Richter (Belgium) and Roche Diagnostics (Switzerland) and for a research cooperation with Ansh Labs (USA). BM reports consultancy for OvsEva, Merck and Guerbet<br>The study was presented at ESHRE 2016 (O-035, O-036, O-037)<br>Email correspondence undertaken: yes, with Helen Torrance (review author)(Torrance 2017 [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| **Bias** | **Authors' judgement** | **Support for judgement** |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Stated as randomisation "using a web-based randomization program" |
| Allocation concealment (selection bias) | Low risk | Not stated in paper, but authors confirmed via email that the trial allocation was only revealed after entering information regarding the participants' eligibility criteria on the web-based system and then clicking 'randomise' |
| Blinding of participants and personnel (performance bias)<br>All outcomes | High risk | Described as 'open-label', and the paper acknowledges the potential for this bias in the observation that there were more cancellations in the 150 IU group for not fulfilling the HCG criterion than in the 450 IU group, and that it is hypothetically possible that standard dosing would have been su- |

**Van Tilborg 2017** (*Continued*)

| | | |
|---|---|---|
| | | perior to increased dosing in women with an AFC 0-7 |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | High risk | Stated as 'open-label' |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | Loss to follow-up was minimal for the first IVF cycle outcomes; however, approximately 20% loss to follow-up for 18-month outcome |
| Selective reporting (reporting bias) | Low risk | Study protocol published and all outcomes reported |
| Other bias | Low risk | None |

**YongPYK 2003**

| | |
|---|---|
| Methods | RCT, **direct dose comparison study**, 123 randomised<br>Study arms: 2<br>Setting: UK, 1 centre: Royal Infirmary of Edinburgh<br>Recruitment period: September 1999 to December 2000 |
| Participants | Population: anticipated **NORMAL** responders<br>Age: eligibility criteria: 23-41 years. Mean (SD) G1: 33.5 years (3.7), G2: 34.2 years (3.3)<br>**Ovarian reserve test: eligibility criteria and demographics**<br>**bFSH (IU/L)** < 10 IU/L, and no previous poor response to stimulation (i.e. 4 oocytes retrieved) or OHSS<br>**AMH**: no eligibility criteria, not recorded<br>**AFC**: no eligibility criteria, not recorded<br>**bFSH (IU/L)**: eligibility criteria < 10 IU/L. Mean (SD) G1: 6.4 IU/L (1.5), G2: 6.9 IU/L (1.6)<br>**Criteria relating to prior IVF/ICSI response**: previous poor response to stimulation (i.e. 4 oocytes retrieved) or OHSS<br>Additional inclusion criteria: BMI < 34 kg/$m^2$, regular menstrual cycles (25-35 days)<br>Additional exclusion criteria: PCOS, 1 ovary or previous ovarian surgery, any chronic cardiovascular, renal, hepatic, or pulmonary disease, oocyte donation cycles |
| Interventions | Group 1 dose/drug: **150 IU** (Gonal-F, Serono)<br>Group 2 dose/drug: **225 IU** (Gonal-F, Serono)<br>Protocol: long agonist<br>Dose titration: not permitted |
| Outcomes | Reported in paper: clinical pregnancy, multiple pregnancy, moderate and severe OHSS, number of oocytes, cycle cancellations for poor and hyper-response, had at least 1 transferable embryo, total dose of FSH, duration of FSH<br>Obtained from author correspondence: none |

**YongPYK 2003** (*Continued*)

| | |
|---|---|
| | Not available: live birth, ongoing pregnancy, poor response, normal response, hyper-response |
| Notes | Trial registration: not registered*<br>Funding: funded by the IVF programme*<br>Conflict of interest: none declared<br>No conference presentation*<br>Email correspondence undertaken: yes with corresponding author KJ Thong (Thong 2017 [pers comm])<br>*indicates information obtained from email correspondence |

***Risk of bias***

| **Bias** | **Authors' judgement** | **Support for judgement** |
|---|---|---|
| Random sequence generation (selection bias) | Low risk | Paper states "Envelopes containing equal numbers of instructions for each treatment had been thoroughly mixed and then numbered consecutively before commencement of the study" |
| Allocation concealment (selection bias) | Low risk | Authors confirmed envelopes met SNOSE criteria |
| Blinding of participants and personnel (performance bias)<br>All outcomes | High risk | Study was not blinded: "The study was subsequently performed in a nonblinded fashion" |
| Blinding of outcome assessment (detection bias): OHSS<br>OHSS | High risk | Study was not blinded: "The study was subsequently performed in a nonblinded fashion" |
| Incomplete outcome data (attrition bias)<br>All outcomes | Low risk | Paper describes 1 person as being excluded; however, it is not clear whether or not this was after randomisation. This remains low risk either way |
| Selective reporting (reporting bias) | High risk | Study not registered and no protocol available*, not reporting live birth |
| Other bias | Low risk | - |

**AFC**: antral follicle count; **AMH**: Anti-Müllerian hormone; **ASRM**: American Society for Reproductive Medicine; **(b)FSH**: (basal) follicle stimulating hormone; **BMI**: body mass index; **ESHRE**: European Society of Human Reproduction and Embryology; **G1/G2/etc.**: group 1/group 2/etc.; **ICSI**: intracytoplasmic sperm injection; **IU**: international units;**IVF**: in vitro fertilisation; **LH**: luteinising hormone; **OHSS**: ovarian hyperstimulation syndrome; **ORT**: ovarian reserve test; **PCOS**: polycystic ovarian syndrome; **RCT**: randomised controlled trial; **SD**: standard deviation; **SNOSE**: sequentially numbered, opaque sealed envelopes.

## Characteristics of excluded studies *[ordered by study ID]*

| Study | Reason for exclusion |
|---|---|
| Berkkanoglu 2010 | Through author correspondence, we discovered the study was not truly randomised: "patients were randomised according to the last number of their patient number; 1,4,7; 2,5,8; and 3,6,9, into the 3 groups"(Berkkanoglu 2017 [pers comm]) |
| Camier 1999 | Did not use an ORT as part of eligibility assessment or report ORT demographics |
| Cavagna 2009 | Did not use an ORT as part of eligibility assessment or report ORT demographics |
| DeJong 2000 | Did not use an ORT as part of eligibility assessment or report ORT demographics |
| Fluker 2000 | Did not use an ORT as part of eligibility assessment or report ORT demographics. Paper states included participants had 'normal day 3 FSH' but no elaboration. Author correspondence undertaken; however, author unable to provide further details. It also appears that this data was combined from 2 individual RCTs |
| Latin-American Puregon IVF Study Group 2001 | Did not use an ORT as part of eligibility assessment or report ORT demographics |
| NCT02915900 | Not comparison of interest as intervention involves a blood test to determine FSH dose, with no involvement of any ORT as defined in this review |
| Out 1999 | Did not use an ORT as part of eligibility assessment or report ORT demographics |
| Out 2000 | Did not use an ORT as part of eligibility assessment or report ORT demographics |
| Out 2001 | This study included participants with LH/FSH ratio of 3 or more, which is not a direct measure of bFSH. No other ORT used |
| Pruksananonda 2004 | Did not use an ORT as part of eligibility assessment or report ORT demographics. Author correspondence confirmed the data provided on 'endocrinological parameterse was LH and not bFSH, therefore no relevant ORT available |
| Simberg 2000 | Only interim analysis available; authors did not respond to emails |
| Tsagareishvili 2005 | Russian study was translated into English, upon which we discovered that the study did not measure any ORT |
| Wikland 2001 | Did not use an ORT as part of eligibility assessment or report ORT demographics. Upon author correspondence the author stated that FSH was measured, and was probably below 10 for inclusion; however as this cannot be confirmed, we excluded this trial |

**(b)FSH**: (basal) follicle stimulating hormone; **LH**: luteinising hormone; **ORT**: ovarian reserve test.

## Characteristics of studies awaiting assessment *[ordered by study ID]*

**NCT02309671**

| | |
|---|---|
| Methods | RCT |
| Participants | Inclusion criteria:<br>• Women diagnosed with tubal infertility, unexplained infertility, infertility related to endometriosis stage I/II or with partners diagnosed with male factor infertility<br>• Women eligible for IVF and/or ICSI treatment<br>• Women aged 20-39 years<br>• Women with body mass index (BMI) of 17.5-32.0 kg/$m^2$<br>Exclusion criteria:<br>• Women with polycystic ovary syndrome (PCOS) associated with anovulation, endometriosis stage III/IV<br>• Women with history of recurrent miscarriage<br>• Women with contraindications to controlled ovarian stimulation with gonadotropins<br>• Women with 3 or more controlled ovarian stimulation cycles |
| Interventions | Novel FSH FE 999049 (3 doses) compared to FOLLISTIM 150 IU |
| Outcomes | Primary outcome measures: number of oocytes retrieved<br>Secondary outcome measures:<br>• Number of follicles during stimulation<br>• Size of follicles during stimulation<br>• Endocrine profile measured by circulating levels of hormones<br>• Total IMP dose administered measured from first until last dose<br>• Embryo quality measured by fertilised oocytes and number and quality of embryos and blastocysts during culturing<br>• Successful pregnancy rate<br>• Frequency of adverse events<br>• Intensity of adverse events |
| Notes | NCT02309671<br>The trial authors presented data as a poster at ESHRE 2017; however, only % are supplied, and it is not possible to extract numerators or denominators. The authors were contacted but were not able to provide these data |

**BMI**: body mass index; **ESHRE**: European Society of Human Reproduction and Embryology; **FSH**: follicle stimulating hormone; **ICSI**: intracytoplasmic sperm injection; **IMP**: intramuscular progesterone; **IU**: international units;**IVF**: in vitro fertilisation; **PCOS**: polycystic ovarian syndrome.

## Characteristics of ongoing studies *[ordered by study ID]*

### CTRI/2016/10/007367

| | |
|---|---|
| Trial name or title | Phase III study to evaluate the efficacy and safety of recombinant human FSH of Cadila Healthcare Limited, India as compared to Gonal-F administered subcutaneously in female patients undergoing assisted reproductive technology |
| Methods | RCT |
| Participants | Inclusion criteria:<br>1. Women between 22-38 years of age with regular menstrual cycle of 24-35 days<br>2. Infertile women undergoing controlled ovarian hyperstimulation (COH) for assisted reproductive technology (ART)<br>3. BMI between 18-30 kg/$m^2$ inclusive<br>4. Transvaginal ultrasound documenting the presence of both ovaries without abnormalities and normal uterine adnexa<br>5. Clinically acceptable ranges of basal follicle-stimulating hormone (FSH), luteinising hormone (LH), estradiol (E2) at the time of enrolment<br>6. Antral follicle count (AFC) 8-25 follicles (sum of both ovaries)<br>7. Willing to comply with all the study requirements and procedures<br>8. Normal or clinically insignificant haematology, serum chemistry and urinalysis parameters during screening<br>9. Willing to provide written informed consent<br>Exclusion criteria:<br>1. History of allergy or hypersensitivity reactions to FSH or any other ingredients of the formulation<br>2. Use of any FSH preparation or clomiphene citrate within 60 days of randomisation<br>3. History of $\geq$ 2 succeeding ART retrieval cycles (which includes fresh and frozen embryo transfers before the study cycle without clinical pregnancy<br>4. Presence of polycystic ovaries (PCO)<br>5. Previous history of severe ovarian hyperstimulation syndrome<br>6. Presence of severe endometriosis (ASRM stage 3 or stage 4) and hydrosalpinx<br>7. Presence or history of thrombophlebitis or thromboembolic disorders<br>8. History of extrauterine pregnancy in the previous 3 months<br>9. History of poor response to gonadotropin treatment (defined as fewer than 5 oocytes retrieved in a previous attempt)<br>10. Subjects with clinically significant unstable medical disorders, life-threatening disease, or current malignancies.<br>11. Positive Pap smear at screening<br>12. Combination or hormonal implants $\leq$ 6 months prior to screening<br>13. Positive pregnancy test at screening |
| Interventions | Appears to be comparing recombinant human FSH of Cadila Healthcare Ltd. with the Gonal-F. However if the study has multiple arms of the same drug it may be eligible for inclusion |
| Outcomes | • Number of oocytes retrieves<br>• Biochemical pregnancy rate after 2 weeks of embryo transfer<br>• Total dose of r-hFSH required<br>• Number of days of r-hFSH stimulation<br>• Number and size distribution of follicles at the day of ovulation induction<br>• Percentage of participants with need to increase or lower the dose of r-hFSH |

**CTRI/2016/10/007367** (*Continued*)

| | |
|---|---|
| | • Number of good quality oocytes |
| Starting date | 22 October 2016<br>Status: open to recruitment (19 May 2017) |
| Contact information | A number of emails are provided on the registration page. We made contact with investigators listed on the trial registration; however, they were unable to update us on the status of this study<br>Garden View Corporate House No. 8, Opp. AUDA Garden, Bodakdev, Ahmedabad No. 8, Opp. AUDA Garden, Bodakdev, Ahmedabad Ahmadabad GUJARAT 380054 India |
| Notes | Accessed 15 May 2017: ctri.nic.in/Clinicaltrials/showallp.php?mid1=16431&EncHid=&userName=CTRI/2016/10/007367 |

**EUCT2012-004969-40**

| | |
|---|---|
| Trial name or title | An AMH based individualised controlled ovarian stimulation regiment using Corifollitrophin or graded doses of rFSH vs a standard protocol. A randomised controlled trial |
| Methods | RCT |
| Participants | Inclusion criteria:<br>1. Women with evidence of COS a view IVFeller ICSI (as per trial website; requires clarification)<br>2. First treatment with IVF/ICSI in the department<br>3. Age 25-38 years;<br>4. AMH is between 5-50<br>5. Weight < 75 kg<br>6. Normal menstrual cycle length of 24-35 days, which are presumably ovulatory<br>7. 2 ovaries<br>8. Uterus with expected normal function (e.g. no clinically significant fibroids) documented by ultrasound at screening<br>9. Willing and able to sign the informed consent<br>Exclusion criteria<br>1. History of current PCOS or endometriosis stage III/IV<br>2. History of severe ovarian hyperstimulation syndrome<br>3. Presence of hydrosalpinx by ultrasound<br>4. History of recurrent consecutive miscarriages (> 3)<br>5. FSH > 12 IU/L (in the early follicular phase)<br>6. Contraindications for use of gonadotropins or GnRH analogues<br>7. History of current epilepsy, HIV infection, diabetes or cardiovascular, gastrointestinal, hepatic, renal, or pulmonary disease<br>8. Pregnancy, lactation or contraindication to pregnancy<br>9. Current or previous (last 12 months) abuse of alcohol or drugs<br>10. History of chemotherapy (except gestational reasons) and radiotherapy<br>11. Undiagnosed vaginal bleeding<br>12. Tumours of the ovary, breast, adrenal, pituitary or hypothalamus and malformations of sexual organs incompatible with pregnancy<br>13. Abnormal karyotype of the patient (if karyotype is performed)<br>14. Hypersensitivity to study drug |

**EUCT2012-004969-40** (*Continued*)

| | |
|---|---|
| Interventions | Unclear |
| Outcomes | Primary outcome: an appropriate or an inappropriate number of oocytes. This should be understood in the term of participants in the 2 arms are classified as having an appropriate response (5 - 14 eggs) or inappropriate response (< 5 or > 14 eggs);<br>Secondary outcomes<br>• Fertilisation rate<br>• Number transferred embryos<br>• Number of participants who achieve blastocyst transfer ring<br>• Implantation rate<br>• Duration of stimulation<br>• Luteal phase inconvenience and enlargement of ovaries<br>• Clinical pregnancy (GA weeks 7-8) |
| Starting date | Start date unclear<br>Status: ongoing (19 May 17)<br>Trial authors confirmed the study has recently completed recruitment; however, trial results will not be available for some time |
| Contact information | Professor Anders Nyboe Andersen, Fertility Clinic, Copenhagen University Hospital, Rigshospitalet |
| Notes | www.clinicaltrialsregister.eu/ctr-search/trial/2012-004969-40/DK<br>Last accessed 24 November 2017 |

**NCT01794208**

| | |
|---|---|
| Trial name or title | Efficacy and safety of FSH-GEX in comparison with 150 IU Gonal-f |
| Methods | RCT |
| Participants | Inclusion criteria:<br>1. Female patient for whom ICSI treatment is justified<br>2. Serum follicle-stimulating hormone concentration<br>3. Anti-mullerian hormone concentration<br>4. Antral follicle count<br>5. Body mass index and body weight<br>6. Presence of both ovaries<br>7. Regular spontaneous cycles between 21 and 35 days in length<br>8. Normal uterine cavity as assessed by transvaginal sonography at screening<br>9. Willing and able to comply with the protocol<br>10. Willing and able to provide written informed consent<br>Exclusion criteria:<br>1. Women who had more than 2 unsuccessful previous assisted reproduction technology cycles before inclusion into the study<br>2. Previous poor responders<br>3. Women with previous hyperstimulation syndrome or cycle cancellation because of imminent hyperstimulation syndrome<br>4. Women with a history of or current polycystic ovarian morphology syndrome |

**NCT01794208** (*Continued*)

| | |
|---|---|
| | 5. Women with a history of or current endometriosis III or IV<br>6. Presence of ovarian cyst at screening<br>7. Any contraindication to becoming pregnant<br>8. History of ≥ 3 clinical or preclinical miscarriages<br>9. Abnormal cervical smear, Papanicolaou (PAP) score ≥ 3<br>10. Any history of malignant cancer other than in situ breast or skin cancer requiring local excision<br>11. Any endocrine abnormalities requiring treatment<br>12. Any clinically significant systematic disease<br>13. Any known infection with human immunodeficiency virus, hepatitis B or C<br>14. History of thrombosis or other risk factors including any coagulation abnormality leading to an increased risk of clotting<br>15. Family history of genetic risk factors concerning pregnancy or birth<br>16. Use of concomitant medication, which in the opinion of the investigator might interfere with ICSI preparation procedures<br>17. Active smoking<br>18. Any active substance abuse of drugs, medications or alcohol within the last 5 years<br>19. Women in an institution by official or court order<br>20. Women who are unable or unwilling to provide informed consent<br>21. Any participation in another clinical trial within the last 60 days before randomisation<br>22. Previous FSH-GEX administration<br>23. Known hypersensitivity to any component of the investigational and non-investigational products used in this study |
| Interventions | Drug: FSH-GEX - 5 arms<br>Drug: Gonal-F (this arm not eligible for this review) |
| Outcomes | Primary outcome measures: number of follicles<br>Secondary outcome measures: pharmacodynamic effect of FSH following administration by subcutaneous injection, follicular response as determined by transvaginal ultrasonography, number of retrieved cumulus-oocyte-complexes, number of oocytes retrieved, number of 2 pronuclei oocytes, biochemical pregnancy rate, rate of clinical pregnancy, implantation rate, pharmacodynamic effect of FSH following administration by subcutaneous injection, estradiol and inhibin B serum levels, ongoing pregnancy rate and live birth rate, incidence of adverse events, ovarian hyperstimulation syndrome, anti-drug-antibodies, overall tolerability |
| Starting date | January 2013<br>Status: this study has been completed (19 May 17) |
| Contact information | Glycotope GmbH (Germany and Hungary); no contact details available |
| Notes | Emails were sent to info@glycotope.com with no response<br>Appears to also have this registration number: EUCTR2012-003006-27-HU |

**NCT02430740**

| | |
|---|---|
| Trial name or title | Tailored ovarian stimulation based on BMI, AMH, AFC |
| Methods | RCT |
| Participants | Inclusion criteria:<br>1. Female infertile women eligible for IVF treatment<br>Exclusion criteria:<br>1. Polycystic ovaries<br>2. Untreated thyroid pathology<br>3. Hypogonadotropic hypogonadism<br>4. UntreaTed hyperprolactinaemia<br>5. Study drug hypersensitivity<br>6. Previous OHSS<br>7. Unilateral ovariectomy<br>8. Genital malformation<br>9. BMI > 40 kg/$m^2$ |
| Interventions | Control group: standard care rFSH.<br>Study group: modified dose of rFSH based on AFC with a correction factor based on BMI and basal AMH level |
| Outcomes | Primary outcomes: number of mature follicles and eggs collected at egg retrieval; amount of rFSH used<br>Secondary outcomes: fertilisation rate; cleavage rate; clinical pregnancy rate; inhibin B and AMH levels during ovarian stimulation |
| Starting date | January 2016<br>Status: this study is currently recruiting participants (19 May 17) |
| Contact information | Christine Wyns, MD, PhD 003227649501 christine.wyns@uclouvain.be<br>Céline Pirard, Md, PhD 003227644116 celine.pirard@uclouvain.be<br>Cliniques universitaires Saint-Luc- Université Catholique de Louvain |
| Notes | Authors contacted who confirmed the study is ongoing |

**NCT02739269**

| | |
|---|---|
| Trial name or title | Antimullerian hormone vs antral follicle count for determination of gonadotrophin dosing in IVF |
| Methods | RCT |
| Participants | Inclusion criteria:<br>1. Women undergoing the first IVF cycle during the study period<br>Exclusion criteria:<br>1. Body mass index ≥ 30 kg/m2<br>2. Women in repeated IVF cycles<br>3. Women undergoing IVF treatment using donor oocytes<br>4. Women undergoing pre-implantation genetic diagnosis |

**NCT02739269** (*Continued*)

| | |
|---|---|
| Interventions | AFC group: starting dose of gonadotrophin will be determined based on serum AMH concentration as follows: AFC ≤ 5: 300 IU daily; AFC 6-15: 225 IU daily; AFC > 15: 150 IU daily<br>AMH group: starting dose of gonadotrophin will be determined based on serum AMH concentration as follows: AMH ≤ 1.0 ng/mL: 300 IU daily; AMH 1.1-3.3 ng/mL: 225 IU daily; AMH > 3.3 ng/mL: 150 IU daily |
| Outcomes | Primary outcome measures: percentage of participants having appropriate ovarian response, Percentage of participants with number of oocytes retrieved being 6-15<br>Secondary outcome measures: percentage of participants requiring step-up or step-down of gonadotrophin dose upon first ultrasound tracking. The dose of gonadotrophin will be adjusted according to the ovarian response: if 5 or fewer follicles growing beyond 10 mm → step up; if more than 15 follicles growing beyond 10 mm → step down |
| Starting date | April 2016<br>Status: this study is currently recruiting participants (19 May 2017) |
| Contact information | Hang Wun Raymond Li, MBBS, FRCOG +852 22553914 raymondli@hku.hk<br>Ernest Hung Yu Ng, MD, FRCOG +852 22553400 nghye@hku.hk<br>The University of Hong Kong |
| Notes | Author correspondence confirmed the study is ongoing with 90/200 subjects recruited (13 April 17) |

**Singh 2015**

| | |
|---|---|
| Trial name or title | A prospective randomised controlled trial (RCT) on the role of AMH tailored stimulation protocols (agonist or antagonist), in improving IVF outcome in previous failed cycles |
| Methods | RCT |
| Participants | 286 currently (recruitment ongoing as per author correspondence) |
| Interventions | AMH-tailored vs untailored dosing |
| Outcomes | Primary outcomes: implantation rate, cumulative pregnancy rate and total cost of cycle<br>Secondary outcomes were cancellation rates and OHSS rates |
| Starting date | 2010 |
| Contact information | Dr Randhir Singh and Dr Monica Singh<br>Bhopal Test-tube-baby and Endoscopy Centre<br>Emails: bttbcentre@gmail.com and iibhopal@gmail.com<br>Cell 09200002833, 9303133385 |
| Notes | Authors provided data in addition to abstract; however, we discovered that study recruitment is ongoing, so we do not include data in this review |

**AFC**: antral follicle count; **AMH**: Anti-Müllerian hormone; **ASRM**: American Society for Reproductive Medicine; **BMI**: body mass index; **GA**: gestational age; **ICSI**: intracytoplasmic sperm injection; **IU**: international units;**OHSS**: ovarian hyperstimulation syndrome; **PCO**: polycystic ovaries;**(r/r-h)FSH**: (recombinant/recombinant-human) follicle stimulating hormone; **RCT**: randomised controlled trial.

## DATA AND ANALYSES

**Comparison 1. Anticipated low responders: higher vs lower dose**

| Outcome or subgroup title | No. of studies | No. of participants | Statistical method | Effect size |
|---|---|---|---|---|
| 1 Live birth or ongoing pregnancy | 4 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 1.1 300/450 IU vs 150 IU | 2 | 286 | Odds Ratio (M-H, Fixed, 95% CI) | 0.71 [0.32, 1.58] |
| 1.2 400/450 IU vs 300 IU | 1 | 62 | Odds Ratio (M-H, Fixed, 95% CI) | 0.77 [0.19, 3.19] |
| 1.3 600 IU vs 450 IU | 1 | 356 | Odds Ratio (M-H, Fixed, 95% CI) | 1.33 [0.71, 2.52] |
| 2 Severe OHSS | 4 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 2.1 300/450 IU vs 150 IU | 2 | 286 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 2.2 400/450 IU vs 300 IU | 1 | 62 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 2.3 600 IU vs 450 IU | 1 | 356 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 3 Clinical pregnancy | 5 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 3.1 300/450 IU vs 150 IU | 2 | 286 | Odds Ratio (M-H, Fixed, 95% CI) | 0.50 [0.25, 1.00] |
| 3.2 400/450 IU vs 300 IU | 2 | 110 | Odds Ratio (M-H, Fixed, 95% CI) | 0.84 [0.26, 2.69] |
| 3.3 600 IU vs 450 IU | 1 | 356 | Odds Ratio (M-H, Fixed, 95% CI) | 1.14 [0.66, 1.99] |
| 4 Time to clinical pregnancy | 0 | 0 | Mean Difference (IV, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 5 Moderate or severe OHSS | 4 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 5.1 300/450 IU vs 150 IU | 2 | 286 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 5.2 450 IU vs 300 IU | 1 | 62 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 5.3 600 IU vs 450 IU | 1 | 356 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 7.23 [0.14, 364.29] |
| 6 Multiple pregnancy in randomised women | 4 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 6.1 300/450 IU vs 150 IU | 2 | 286 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.55 [0.06, 5.31] |
| 6.2 450 IU vs 300 IU | 1 | 62 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 6.3 600 IU vs 450 IU | 1 | 356 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.49 [0.16, 1.55] |
| 7 Log(N oocytes retrieved); data presented on the log scale - cannot interpret pooled estimates as N of oocytes | 5 | | Mean Difference (Fixed, 95% CI) | Subtotals only |
| 7.1 300/450 IU vs 150 IU | 2 | | Mean Difference (Fixed, 95% CI) | 0.69 [0.50, 0.88] |
| 7.2 400/450 IU vs 300 IU | 2 | | Mean Difference (Fixed, 95% CI) | -0.03 [-0.30, 0.24] |
| 7.3 600 IU vs 450 IU | 1 | | Mean Difference (Fixed, 95% CI) | 0.08 [-0.04, 0.20] |
| 8 Poor response to stimulation | 2 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 8.1 300/450 IU vs 150 IU | 2 | 286 | Odds Ratio (M-H, Fixed, 95% CI) | 0.52 [0.32, 0.84] |
| 9 Normal response to stimulation | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 9.1 300/450 IU vs 150 IU | 1 | 234 | Odds Ratio (M-H, Fixed, 95% CI) | 1.79 [1.05, 3.04] |
| 10 Hyper-response to stimulation | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 10.1 300/450 IU vs 150 IU | 1 | 234 | Odds Ratio (M-H, Fixed, 95% CI) | 4.53 [0.94, 21.82] |
| 11 Cycle cancellations for poor response | 5 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 11.1 300/450 IU vs 150 IU | 2 | 286 | Odds Ratio (M-H, Fixed, 95% CI) | 0.23 [0.11, 0.47] |
| 11.2 400/450 IU vs 300 IU | 2 | 110 | Odds Ratio (M-H, Fixed, 95% CI) | 1.47 [0.62, 3.49] |
| 11.3 600 IU vs 450 IU | 1 | 356 | Odds Ratio (M-H, Fixed, 95% CI) | 0.86 [0.50, 1.50] |
| 12 Cycle cancellations for hyper-response | 5 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 12.1 300/450 IU vs 150 IU | 2 | 286 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 7.93 [0.16, 400.62] |

| | | | | |
|---|---|---|---|---|
| 12.2 400/450 IU vs 300 IU | 2 | 110 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 12.3 600 IU vs 450 IU | 1 | 356 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 13 Cycle cancellations for poor or hyper-response | 5 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 13.1 300/450 IU vs 150 IU | 2 | 286 | Odds Ratio (M-H, Fixed, 95% CI) | 0.25 [0.13, 0.50] |
| 13.2 400/450 IU vs 300 IU | 2 | 110 | Odds Ratio (M-H, Fixed, 95% CI) | 1.47 [0.62, 3.49] |
| 13.3 600 IU vs 450 IU | 1 | 356 | Odds Ratio (M-H, Fixed, 95% CI) | 0.86 [0.50, 1.50] |
| 14 Women with at least one transferable embryo | 5 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 14.1 300/450 IU vs 150 IU | 2 | 286 | Odds Ratio (M-H, Fixed, 95% CI) | 1.76 [1.07, 2.87] |
| 14.2 400/450 IU vs 300 IU | 2 | 110 | Odds Ratio (M-H, Fixed, 95% CI) | 0.71 [0.31, 1.60] |
| 14.3 600 IU vs 450 IU | 1 | 356 | Odds Ratio (M-H, Fixed, 95% CI) | 1.19 [0.78, 1.82] |
| 15 Total dose of FSH | 5 | | Mean Difference (IV, Fixed, 95% CI) | Subtotals only |
| 15.1 300/450 IU vs 150 IU | 2 | 286 | Mean Difference (IV, Fixed, 95% CI) | 2.78 [2.57, 3.00] |
| 15.2 400/450 IU vs 300 IU | 2 | 110 | Mean Difference (IV, Fixed, 95% CI) | 1.11 [0.91, 1.31] |
| 15.3 600 IU vs 450 IU | 1 | 356 | Mean Difference (IV, Fixed, 95% CI) | 1.20 [1.07, 1.33] |
| 16 Duration of FSH administration | 4 | | Mean Difference (IV, Fixed, 95% CI) | Subtotals only |
| 16.1 300/450 IU vs 150 IU | 1 | 234 | Mean Difference (IV, Fixed, 95% CI) | -0.70 [-1.48, 0.08] |
| 16.2 400/450 IU vs 300 IU | 2 | 110 | Mean Difference (IV, Fixed, 95% CI) | -0.67 [-1.39, 0.06] |
| 16.3 600 IU vs 450 IU | 1 | 356 | Mean Difference (IV, Fixed, 95% CI) | -1.0 [-1.27, -0.73] |
| 17 Cost per woman randomised | 0 | 0 | Mean Difference (IV, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 18 Cumulative live birth: 1 cycle (fresh + frozen) | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 18.1 300/450 IU vs 150 IU | 1 | 234 | Odds Ratio (M-H, Fixed, 95% CI) | 0.78 [0.35, 1.73] |
| 19 Cumulative live birth: 18 months | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 19.1 300/450 IU vs 150 IU | 1 | 234 | Odds Ratio (M-H, Fixed, 95% CI) | 0.78 [0.46, 1.32] |
| 20 Multiple pregnancy in women with clinical pregnancy | 4 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 20.1 300/450 IU vs 150 IU | 2 | 41 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.92 [0.08, 10.54] |
| 20.2 400/450 IU vs 300 IU | 1 | 9 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 20.3 600 IU vs 450 IU | 1 | 60 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.37 [0.11, 1.31] |

**Comparison 2. Anticipated normal responders: higher vs lower dose**

| Outcome or subgroup title | No. of studies | No. of participants | Statistical method | Effect size |
|---|---|---|---|---|
| 1 Live birth or ongoing pregnancy | 4 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 1.1 200 IU vs 100 IU | 2 | 522 | Odds Ratio (M-H, Fixed, 95% CI) | 0.88 [0.57, 1.36] |
| 1.2 225/200 IU vs 150 IU | 1 | 277 | Odds Ratio (M-H, Fixed, 95% CI) | 1.03 [0.57, 1.86] |
| 1.3 300 IU vs 225 IU | 1 | 135 | Odds Ratio (M-H, Fixed, 95% CI) | 0.65 [0.32, 1.32] |
| 2 Severe OHSS | 7 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 2.1 200 IU vs 100 IU | 2 | 522 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.14 [0.00, 6.96] |
| 2.2 225/200 IU vs 150 IU | 4 | 740 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 1.00 [0.20, 5.02] |
| 2.3 300 IU vs 225 IU | 1 | 135 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.14 [0.00, 6.92] |
| 3 Clinical pregnancy | 7 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 3.1 200 IU vs 100 IU | 1 | 330 | Odds Ratio (M-H, Fixed, 95% CI) | 0.86 [0.50, 1.49] |

| | | | | |
|---|---|---|---|---|
| 3.2 225/200 IU vs 150 IU | 5 | 1037 | Odds Ratio (M-H, Fixed, 95% CI) | 0.98 [0.73, 1.31] |
| 3.3 300 IU vs 225 IU | 1 | 135 | Odds Ratio (M-H, Fixed, 95% CI) | 0.91 [0.46, 1.80] |
| 4 Time to clinical pregnancy | 0 | 0 | Mean Difference (IV, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 5 Moderate or severe OHSS | 7 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 5.1 200 IU vs 100 IU | 2 | 522 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.62 [0.21, 1.87] |
| 5.2 225/200 IU vs 150 IU | 4 | 740 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 1.21 [0.51, 2.85] |
| 5.3 300 IU vs 225 IU | 1 | 135 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.67 [0.11, 3.99] |
| 6 Multiple pregnancy in randomised women | 4 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 6.1 200 IU vs 100 IU | 1 | 330 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.97 [0.38, 2.52] |
| 6.2 225/200 IU vs 150 IU | 2 | 400 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 1.91 [0.38, 9.69] |
| 6.3 300 IU vs 225 IU | 1 | 135 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 7.61 [0.47, 123.02] |
| 7 Log(N oocytes retrieved); data presented on the log scale - cannot interpret pooled estimates as N of oocytes | 8 | | Mean Difference (Fixed, 95% CI) | Subtotals only |
| 7.1 200 IU vs 100 IU | 2 | | Mean Difference (Fixed, 95% CI) | 0.46 [0.36, 0.57] |
| 7.2 225/200 IU vs 150 IU | 5 | | Mean Difference (Fixed, 95% CI) | 0.16 [0.08, 0.24] |
| 7.3 300 IU vs 225 IU | 1 | | Mean Difference (Fixed, 95% CI) | 0.03 [-0.17, 0.23] |
| 8 Poor response to stimulation | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 8.1 225/200 IU vs 150 IU | 1 | 277 | Odds Ratio (M-H, Fixed, 95% CI) | 0.50 [0.30, 0.83] |
| 9 Normal response to stimulation | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 9.1 225/200 IU vs 150 IU | 1 | 277 | Odds Ratio (M-H, Fixed, 95% CI) | 1.26 [0.78, 2.04] |
| 10 Hyper-response to stimulation | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 10.1 225/200 IU vs 150 IU | 1 | 277 | Odds Ratio (M-H, Fixed, 95% CI) | 4.08 [1.47, 11.34] |
| 11 Cycle cancellations for poor response | 8 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 11.1 200 IU vs 100 IU | 2 | 522 | Odds Ratio (M-H, Fixed, 95% CI) | 0.33 [0.16, 0.66] |
| 11.2 225/200 IU vs 150 IU | 5 | 1037 | Odds Ratio (M-H, Fixed, 95% CI) | 0.56 [0.36, 0.88] |
| 11.3 300 IU vs 225 IU | 1 | 135 | Odds Ratio (M-H, Fixed, 95% CI) | 0.11 [0.01, 2.01] |
| 12 Cycle cancellations for hyper-response | 8 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 12.1 200 IU vs 100 IU | 2 | 522 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 1.93 [0.20, 18.62] |
| 12.2 225/200 IU vs 150 IU | 5 | 1037 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 2.28 [0.99, 5.26] |
| 12.3 300 IU vs 225 IU | 1 | 135 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 1.02 [0.06, 16.40] |
| 13 Cycle cancellations for poor or hyper-response | 8 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 13.1 200 IU vs 100 IU | 2 | 522 | Odds Ratio (M-H, Fixed, 95% CI) | 0.37 [0.19, 0.72] |
| 13.2 225/200 IU vs 150 IU | 5 | 1037 | Odds Ratio (M-H, Fixed, 95% CI) | 0.76 [0.51, 1.13] |
| 13.3 300 IU vs 225 IU | 1 | 135 | Odds Ratio (M-H, Fixed, 95% CI) | 0.19 [0.02, 1.68] |
| 14 Women with at least one transferable embryo | 8 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 14.1 200 IU vs 100 IU | 2 | 522 | Odds Ratio (M-H, Fixed, 95% CI) | 1.58 [0.95, 2.64] |
| 14.2 225/200 IU vs 150 IU | 5 | 1037 | Odds Ratio (M-H, Fixed, 95% CI) | 1.06 [0.76, 1.47] |
| 14.3 300 IU vs 225 IU | 1 | 135 | Odds Ratio (M-H, Fixed, 95% CI) | 1.75 [0.60, 5.13] |
| 15 Total dose of FSH | 8 | | Mean Difference (IV, Fixed, 95% CI) | Subtotals only |
| 15.1 200 IU vs 100 IU | 2 | 522 | Mean Difference (IV, Fixed, 95% CI) | 795.79 [656.67, 934.91] |
| 15.2 225/200 IU vs 150 IU | 5 | 1037 | Mean Difference (IV, Fixed, 95% CI) | 503.12 [456.23, 550.00] |
| 15.3 300 IU vs 225 IU | 1 | 135 | Mean Difference (IV, Fixed, 95% CI) | 725.0 [597.44, 852.56] |

| | | | | |
|---|---|---|---|---|
| 16 Duration of FSH administration | 6 | | Mean Difference (IV, Fixed, 95% CI) | Subtotals only |
| 16.1 200 IU vs 100 IU | 1 | 330 | Mean Difference (IV, Fixed, 95% CI) | -1.80 [-2.21, -1.39] |
| 16.2 225/200 IU vs 150 IU | 4 | 961 | Mean Difference (IV, Fixed, 95% CI) | -0.25 [-0.51, 0.01] |
| 16.3 300 IU vs 225 IU | 1 | 135 | Mean Difference (IV, Fixed, 95% CI) | -0.30 [-0.79, 0.19] |
| 17 Cost per woman randomised | 0 | 0 | Mean Difference (IV, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 18 Cumulative live birth: 1 cycle (fresh + frozen) | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 18.1 225/200 IU vs 150 IU | 1 | 277 | Odds Ratio (M-H, Fixed, 95% CI) | 0.88 [0.51, 1.52] |
| 19 Cumulative live birth: 18 months | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 19.1 225/200 IU vs 150 IU | 1 | 277 | Odds Ratio (M-H, Fixed, 95% CI) | 1.01 [0.63, 1.62] |
| 20 Multiple pregnancy in women with clinical pregnancy | 4 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 20.1 200 IU vs 100 IU | 1 | 63 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 1.14 [0.39, 3.38] |
| 20.2 225/200 IU vs 150 IU | 2 | 89 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 1.67 [0.28, 9.76] |
| 20.3 300 IU vs 225 IU | 1 | 58 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 8.24 [0.50, 135.17] |
| 21 Dose-response: live birth or ongoing pregnancy | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 21.1 12.1$\mu$g vs 10.3$\mu$g | 1 | 41 | Odds Ratio (M-H, Fixed, 95% CI) | 1.5 [0.39, 5.84] |
| 21.2 10.3$\mu$g vs 8.6$\mu$g | 1 | 39 | Odds Ratio (M-H, Fixed, 95% CI) | 0.92 [0.25, 3.42] |
| 21.3 8.6$\mu$g vs 6.9$\mu$g | 1 | 40 | Odds Ratio (M-H, Fixed, 95% CI) | 0.62 [0.16, 2.43] |
| 21.4 6.9$\mu$g vs 5.2$\mu$g | 1 | 38 | Odds Ratio (M-H, Fixed, 95% CI) | 1.0 [0.25, 3.93] |
| 22 Dose-response: cumulative live birth: 1 cycle (fresh + frozen) | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 22.1 12.1$\mu$g vs 10.3$\mu$g | 1 | 41 | Odds Ratio (M-H, Fixed, 95% CI) | 1.14 [0.32, 4.08] |
| 22.2 10.3$\mu$g vs 8.6$\mu$g | 1 | 40 | Odds Ratio (M-H, Fixed, 95% CI) | 0.81 [0.22, 2.91] |
| 22.3 8.6$\mu$g vs 6.9$\mu$g | 1 | 39 | Odds Ratio (M-H, Fixed, 95% CI) | 0.73 [0.19, 2.79] |
| 22.4 6.9$\mu$g vs 5.2$\mu$g | 1 | 38 | Odds Ratio (M-H, Fixed, 95% CI) | 1.58 [0.42, 5.95] |
| 23 Dose-response: clinical pregnancy | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 23.1 12.1$\mu$g vs 10.3$\mu$g | 1 | 41 | Odds Ratio (M-H, Fixed, 95% CI) | 1.2 [0.30, 4.80] |
| 23.2 10.3$\mu$g vs 8.6$\mu$g | 1 | 40 | Odds Ratio (M-H, Fixed, 95% CI) | 0.78 [0.19, 3.13] |
| 23.3 8.6$\mu$g vs 6.9$\mu$g | 1 | 39 | Odds Ratio (M-H, Fixed, 95% CI) | 0.92 [0.25, 3.42] |
| 23.4 6.9$\mu$g vs 5.2$\mu$g | 1 | 38 | Odds Ratio (M-H, Fixed, 95% CI) | 1.0 [0.25, 3.93] |
| 24 Dose-response: log(N oocytes retrieved); data presented on the log scale - cannot interpret pooled estimates as N of oocytes | 1 | | Mean Difference (Fixed, 95% CI) | Subtotals only |
| 24.1 12.1$\mu$g vs 10.3$\mu$g | 1 | | Mean Difference (Fixed, 95% CI) | 0.26 [-0.05, 0.57] |
| 24.2 10.3$\mu$g vs 8.6$\mu$g | 1 | | Mean Difference (Fixed, 95% CI) | 0.01 [-0.31, 0.32] |
| 24.3 8.6$\mu$g vs 6.9$\mu$g | 1 | | Mean Difference (Fixed, 95% CI) | 0.28 [-0.13, 0.69] |
| 24.4 6.9$\mu$g vs 5.2$\mu$g | 1 | | Mean Difference (Fixed, 95% CI) | 0.29 [-0.12, 0.70] |
| 25 Dose-response: poor response to stimulation | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 25.1 12.1$\mu$g vs 10.3$\mu$g | 1 | 41 | Odds Ratio (M-H, Fixed, 95% CI) | 0.95 [0.12, 7.46] |
| 25.2 10.3$\mu$g vs 8.6$\mu$g | 1 | 40 | Odds Ratio (M-H, Fixed, 95% CI) | 0.63 [0.09, 4.24] |
| 25.3 8.6$\mu$g vs 6.9$\mu$g | 1 | 39 | Odds Ratio (M-H, Fixed, 95% CI) | 0.49 [0.10, 2.44] |
| 25.4 6.9$\mu$g vs 5.2$\mu$g | 1 | 38 | Odds Ratio (M-H, Fixed, 95% CI) | 0.77 [0.19, 3.16] |
| 26 Dose-response: normal response to stimulation | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |

| | | | | |
|---|---|---|---|---|
| 26.1 12.1μg vs 10.3μg | 1 | 41 | Odds Ratio (M-H, Fixed, 95% CI) | 0.47 [0.08, 2.92] |
| 26.2 10.3μg vs 8.6μg | 1 | 40 | Odds Ratio (M-H, Fixed, 95% CI) | 6.0 [1.08, 33.27] |
| 26.3 8.6μg vs 6.9μg | 1 | 39 | Odds Ratio (M-H, Fixed, 95% CI) | 2.33 [0.55, 9.83] |
| 26.4 6.9μg vs 5.2μg | 1 | 38 | Odds Ratio (M-H, Fixed, 95% CI) | 0.79 [0.21, 3.03] |
| 27 Dose-response: hyper-response to stimulation | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 27.1 12.1μg vs 10.3μg | 1 | 41 | Odds Ratio (M-H, Fixed, 95% CI) | 5.26 [0.24, 116.57] |
| 27.2 10.3μg vs 8.6μg | 1 | 40 | Odds Ratio (M-H, Fixed, 95% CI) | 0.32 [0.01, 8.26] |
| 27.3 8.6μg vs 6.9μg | 1 | 39 | Odds Ratio (M-H, Fixed, 95% CI) | 0.45 [0.04, 5.39] |
| 27.4 6.9μg vs 5.2μg | 1 | 38 | Odds Ratio (M-H, Fixed, 95% CI) | 5.57 [0.25, 124.19] |
| 28 Dose-response: cycle cancellations for poor response | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 28.1 12.1μg vs 10.3μg | 1 | 41 | Odds Ratio (M-H, Fixed, 95% CI) | 3.0 [0.12, 78.04] |
| 28.2 10.3μg vs 8.6μg | 1 | 40 | Odds Ratio (M-H, Fixed, 95% CI) | 0.32 [0.01, 8.26] |
| 28.3 8.6μg vs 6.9μg | 1 | 39 | Odds Ratio (M-H, Fixed, 95% CI) | 0.95 [0.06, 16.31] |
| 28.4 6.9μg vs 5.2μg | 1 | 38 | Odds Ratio (M-H, Fixed, 95% CI) | 0.47 [0.04, 5.70] |
| 29 Dose-response: women with at least one transferable embryo | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 29.1 12.1μg vs 10.3μg | 1 | 41 | Odds Ratio (M-H, Fixed, 95% CI) | 1.68 [0.25, 11.27] |
| 29.2 10.3μg vs 8.6μg | 1 | 40 | Odds Ratio (M-H, Fixed, 95% CI) | 0.63 [0.09, 4.24] |
| 29.3 8.6μg vs 6.9μg | 1 | 39 | Odds Ratio (M-H, Fixed, 95% CI) | 1.69 [0.25, 11.42] |
| 29.4 6.9μg vs 5.2μg | 1 | 38 | Odds Ratio (M-H, Fixed, 95% CI) | 0.63 [0.09, 4.26] |
| 30 Dose-response: total dose of FSH | 1 | | Mean Difference (IV, Fixed, 95% CI) | Subtotals only |
| 30.1 12.1μg vs 10.3μg | 1 | 41 | Mean Difference (IV, Fixed, 95% CI) | 14.20 [0.28, 28.12] |
| 30.2 10.3μg vs 8.6μg | 1 | 40 | Mean Difference (IV, Fixed, 95% CI) | 8.20 [-0.14, 16.54] |
| 30.3 8.6μg vs 6.9μg | 1 | 39 | Mean Difference (IV, Fixed, 95% CI) | 13.5 [5.83, 21.17] |
| 30.4 6.9μg vs 5.2μg | 1 | 38 | Mean Difference (IV, Fixed, 95% CI) | 11.30 [3.44, 19.16] |
| 31 Dose-response: duration of FSH administration | 1 | | Mean Difference (IV, Fixed, 95% CI) | Subtotals only |
| 31.1 12.1μg vs 10.3μg | 1 | 41 | Mean Difference (IV, Fixed, 95% CI) | 0.0 [-1.08, 1.08] |
| 31.2 10.3μg vs 8.6μg | 1 | 40 | Mean Difference (IV, Fixed, 95% CI) | -0.60 [-1.50, 0.30] |
| 31.3 8.6μg vs 6.9μg | 1 | 39 | Mean Difference (IV, Fixed, 95% CI) | -0.10 [-1.12, 0.92] |
| 31.4 6.9μg vs 5.2μg | 1 | 38 | Mean Difference (IV, Fixed, 95% CI) | -0.60 [-1.91, 0.71] |

**Comparison 3. Anticipated high-responders: higher vs lower dose**

| Outcome or subgroup title | No. of studies | No. of participants | Statistical method | Effect size |
|---|---|---|---|---|
| 1 Live birth or ongoing pregnancy | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 1.1 150 IU vs 100 IU | 1 | 521 | Odds Ratio (M-H, Fixed, 95% CI) | 0.98 [0.66, 1.46] |
| 2 Severe OHSS | 1 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 2.1 150 IU vs 100 IU | 1 | 521 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.72 [0.16, 3.19] |
| 3 Clinical pregnancy | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 3.1 150 IU vs 100 IU | 1 | 521 | Odds Ratio (M-H, Fixed, 95% CI) | 1.14 [0.78, 1.66] |
| 4 Time to clinical pregnancy | 0 | | Mean Difference (IV, Fixed, 95% CI) | Subtotals only |
| 5 Moderate or severe OHSS | 1 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 5.1 150 IU vs 100 IU | 1 | 521 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 2.31 [0.80, 6.67] |

| | | | | |
|---|---|---|---|---|
| 6 Multiple pregnancy in randomised women | 1 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 6.1 150 IU vs 100 IU | 1 | 521 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 1.87 [0.19, 18.09] |
| 7 Log(N oocytes retrieved); data presented on the log scale - cannot interpret pooled estimates as N of oocytes | 1 | | Mean Difference (Fixed, 95% CI) | Subtotals only |
| 7.1 150 IU vs 100 IU | 1 | | Mean Difference (Fixed, 95% CI) | 0.67 [0.55, 0.79] |
| 8 Poor response to stimulation | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 8.1 150 IU vs 100 IU | 1 | 521 | Odds Ratio (M-H, Fixed, 95% CI) | 0.15 [0.09, 0.25] |
| 9 Normal response to stimulation | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 9.1 150 IU vs 100 IU | 1 | 521 | Odds Ratio (M-H, Fixed, 95% CI) | 1.07 [0.76, 1.50] |
| 10 Hyper-response to stimulation | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 10.1 150 IU vs 100 IU | 1 | 521 | Odds Ratio (M-H, Fixed, 95% CI) | 5.04 [3.17, 8.02] |
| 11 Cycle cancellations for poor response | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 11.1 150 IU vs 100 IU | 1 | 521 | Odds Ratio (M-H, Fixed, 95% CI) | 0.13 [0.06, 0.28] |
| 12 Cycle cancellations for hyper-response | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 12.1 150 IU vs 100 IU | 1 | 521 | Odds Ratio (M-H, Fixed, 95% CI) | 5.28 [2.16, 12.90] |
| 13 Cycle cancellations for poor or hyper-response | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 13.1 150 IU vs 100 IU | 1 | 521 | Odds Ratio (M-H, Fixed, 95% CI) | 0.57 [0.36, 0.89] |
| 14 Women with at least one transferable embryo | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 14.1 150 IU vs 100 IU | 1 | 521 | Odds Ratio (M-H, Fixed, 95% CI) | 2.33 [1.53, 3.55] |
| 15 Total dose of FSH | 1 | | Mean Difference (IV, Fixed, 95% CI) | Subtotals only |
| 15.1 150 IU vs 100 IU | 1 | 521 | Mean Difference (IV, Fixed, 95% CI) | 345.0 [280.34, 409.66] |
| 16 Duration of FSH administration | 1 | | Mean Difference (IV, Fixed, 95% CI) | Subtotals only |
| 16.1 150 IU vs 100 IU | 1 | 521 | Mean Difference (IV, Fixed, 95% CI) | -1.40 [-1.91, -0.89] |
| 17 Cost per woman randomised | 1 | | Mean Difference (Fixed, 95% CI) | Subtotals only |
| 17.1 150 IU vs 100 IU | 1 | | Mean Difference (Fixed, 95% CI) | -92.0 [-325.24, 141.24] |
| 18 Cumulative live birth: 1 cycle (fresh + frozen) | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 18.1 150 IU vs 100 IU | 1 | 521 | Odds Ratio (M-H, Fixed, 95% CI) | 1.16 [0.81, 1.65] |
| 19 Cumulative live birth: 18 months | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 19.1 150 IU vs 100 IU | 1 | 521 | Odds Ratio (M-H, Fixed, 95% CI) | 1.16 [0.80, 1.68] |
| 20 Multiple pregnancy in women with clinical pregnancy | 1 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 20.1 150 IU vs 100 IU | 1 | 150 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 1.72 [0.18, 16.89] |
| 21 Dose-response: live birth or ongoing pregnancy | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 21.1 12.1$\mu$g vs 10.3$\mu$g | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 2.57 [0.77, 8.57] |
| 21.2 10.3$\mu$g vs 8.6$\mu$g | 1 | 48 | Odds Ratio (M-H, Fixed, 95% CI) | 0.56 [0.16, 1.92] |
| 21.3 8.6$\mu$g vs 6.9$\mu$g | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 0.82 [0.26, 2.55] |
| 21.4 6.9$\mu$g vs 5.2$\mu$g | 1 | 49 | Odds Ratio (M-H, Fixed, 95% CI) | 1.14 [0.36, 3.58] |
| 22 Dose-response: clinical pregnancy | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |

| | | | | |
|---|---|---|---|---|
| 22.1 12.1μg vs 10.3μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 2.57 [0.77, 8.57] |
| 22.2 10.3μg vs 8.6μg | 1 | 48 | Odds Ratio (M-H, Fixed, 95% CI) | 0.56 [0.16, 1.92] |
| 22.3 8.6μg vs 6.9μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 0.7 [0.23, 2.17] |
| 22.4 6.9μg vs 5.2μg | 1 | 49 | Odds Ratio (M-H, Fixed, 95% CI) | 1.33 [0.43, 4.16] |
| 23 Dose-response: cumulative live birth:1 cycle (fresh + frozen) | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 23.1 12.1μg vs 10.3μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 1.67 [0.54, 5.15] |
| 23.2 10.3μg vs 8.6μg | 1 | 48 | Odds Ratio (M-H, Fixed, 95% CI) | 0.71 [0.22, 2.24] |
| 23.3 8.6μg vs 6.9μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 0.73 [0.24, 2.21] |
| 23.4 6.9μg vs 5.2μg | 1 | 49 | Odds Ratio (M-H, Fixed, 95% CI) | 1.52 [0.49, 4.69] |
| 24 Dose-response: severe OHSS | 1 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 24.1 12.1μg vs 10.3μg | 1 | 50 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 7.12 [0.43, 117.44] |
| 24.2 10.3μg vs 8.6μg | 1 | 48 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.14 [0.00, 6.82] |
| 24.3 8.6μg vs 6.9μg | 1 | 50 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 8.03 [0.16, 406.02] |
| 24.4 6.9μg vs 5.2μg | 1 | 49 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 25 Dose-response: moderate or severe OHSS | 1 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 25.1 12.1μg vs 10.3μg | 1 | 50 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 2.67 [0.35, 20.21] |
| 25.2 10.3μg vs 8.6μg | 1 | 48 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 1.0 [0.06, 16.47] |
| 25.3 8.6μg vs 6.9μg | 1 | 50 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 8.03 [0.16, 406.02] |
| 25.4 6.9μg vs 5.2μg | 1 | 49 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 26 Dose-response: log(N oocytes retrieved); data presented on the log scale - cannot interpret pooled estimates as N of oocytes | 1 | | Mean Difference (Fixed, 95% CI) | Subtotals only |
| 26.1 12.1μg vs 10.3μg | 1 | | Mean Difference (Fixed, 95% CI) | 0.12 [-0.13, 0.37] |
| 26.2 10.3μg vs 8.6μg | 1 | | Mean Difference (Fixed, 95% CI) | 0.25 [-0.02, 0.52] |
| 26.3 8.6μg vs 6.9μg | 1 | | Mean Difference (Fixed, 95% CI) | 0.21 [-0.10, 0.52] |
| 26.4 6.9μg vs 5.2μg | 1 | | Mean Difference (Fixed, 95% CI) | 0.41 [0.06, 0.76] |
| 27 Dose-response: poor response to stimulation | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 27.1 12.1μg vs 10.3μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 0.17 [0.01, 3.72] |
| 27.2 10.3μg vs 8.6μg | 1 | 48 | Odds Ratio (M-H, Fixed, 95% CI) | 2.09 [0.18, 24.73] |
| 27.3 8.6μg vs 6.9μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 0.24 [0.02, 2.31] |
| 27.4 6.9μg vs 5.2μg | 1 | 49 | Odds Ratio (M-H, Fixed, 95% CI) | 0.28 [0.07, 1.10] |
| 28 Dose-response: normal response to stimulation | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 28.1 12.1μg vs 10.3μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 1.0 [0.33, 3.03] |
| 28.2 10.3μg vs 8.6μg | 1 | 48 | Odds Ratio (M-H, Fixed, 95% CI) | 0.33 [0.10, 1.13] |
| 28.3 8.6μg vs 6.9μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 1.59 [0.47, 5.42] |
| 28.4 6.9μg vs 5.2μg | 1 | 49 | Odds Ratio (M-H, Fixed, 95% CI) | 1.73 [0.55, 5.47] |
| 29 Dose-response: hyper-response to stimulation | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 29.1 12.1μg vs 10.3μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 1.4 [0.46, 4.28] |
| 29.2 10.3μg vs 8.6μg | 1 | 48 | Odds Ratio (M-H, Fixed, 95% CI) | 2.71 [0.76, 9.73] |
| 29.3 8.6μg vs 6.9μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 1.11 [0.28, 4.42] |
| 29.4 6.9μg vs 5.2μg | 1 | 49 | Odds Ratio (M-H, Fixed, 95% CI) | 2.5 [0.44, 14.36] |
| 30 Dose-response: cycle cancellations for poor response | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 30.1 12.1μg vs 10.3μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 30.2 10.3μg vs 8.6μg | 1 | 48 | Odds Ratio (M-H, Fixed, 95% CI) | 0.32 [0.01, 8.25] |

| | | | | |
|---|---|---|---|---|
| 30.3 8.6μg vs 6.9μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 3.38 [0.13, 87.11] |
| 30.4 6.9μg vs 5.2μg | 1 | 49 | Odds Ratio (M-H, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 31 Dose-response: cycle cancellations for hyper-response | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 31.1 12.1μg vs 10.3μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 0.30 [0.01, 7.61] |
| 31.2 10.3μg vs 8.6μg | 1 | 48 | Odds Ratio (M-H, Fixed, 95% CI) | 3.13 [0.12, 80.68] |
| 31.3 8.6μg vs 6.9μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 0.35 [0.01, 8.93] |
| 31.4 6.9μg vs 5.2μg | 1 | 49 | Odds Ratio (M-H, Fixed, 95% CI) | 2.76 [0.11, 71.25] |
| 32 Dose-response: cycle cancellations for poor or hyper-response | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 32.1 12.1μg vs 10.3μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 0.30 [0.01, 7.61] |
| 32.2 10.3μg vs 8.6μg | 1 | 48 | Odds Ratio (M-H, Fixed, 95% CI) | 1.0 [0.06, 16.97] |
| 32.3 8.6μg vs 6.9μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 1.09 [0.06, 18.40] |
| 32.4 6.9μg vs 5.2μg | 1 | 49 | Odds Ratio (M-H, Fixed, 95% CI) | 0.28 [0.01, 7.30] |
| 33 Dose-response: women with at least one transferable embryo | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 33.1 12.1μg vs 10.3μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 6.58 [0.71, 61.08] |
| 33.2 10.3μg vs 8.6μg | 1 | 48 | Odds Ratio (M-H, Fixed, 95% CI) | 0.35 [0.06, 1.99] |
| 33.3 8.6μg vs 6.9μg | 1 | 50 | Odds Ratio (M-H, Fixed, 95% CI) | 1.43 [0.22, 9.42] |
| 33.4 6.9μg vs 5.2μg | 1 | 49 | Odds Ratio (M-H, Fixed, 95% CI) | 0.73 [0.11, 4.81] |
| 34 Dose-response: total dose of FSH | 1 | | Mean Difference (IV, Fixed, 95% CI) | Subtotals only |
| 34.1 12.1μg vs 10.3μg | 1 | 50 | Mean Difference (IV, Fixed, 95% CI) | 18.90 [11.03, 26.77] |
| 34.2 10.3μg vs 8.6μg | 1 | 48 | Mean Difference (IV, Fixed, 95% CI) | 9.80 [1.34, 18.26] |
| 34.3 8.6μg vs 6.9μg | 1 | 50 | Mean Difference (IV, Fixed, 95% CI) | 7.30 [-1.17, 15.77] |
| 34.4 6.9μg vs 5.2μg | 1 | 49 | Mean Difference (IV, Fixed, 95% CI) | 12.20 [5.05, 19.35] |
| 35 Dose-response: duration of FSH administration | 1 | | Mean Difference (IV, Fixed, 95% CI) | Subtotals only |
| 35.1 12.1μg vs 10.3μg | 1 | 50 | Mean Difference (IV, Fixed, 95% CI) | 0.40 [-0.30, 1.10] |
| 35.2 10.3μg vs 8.6μg | 1 | 48 | Mean Difference (IV, Fixed, 95% CI) | -0.40 [-1.32, 0.52] |
| 35.3 8.6μg vs 6.9μg | 1 | 50 | Mean Difference (IV, Fixed, 95% CI) | -1.0 [-2.11, 0.11] |
| 35.4 6.9μg vs 5.2μg | 1 | 49 | Mean Difference (IV, Fixed, 95% CI) | -0.70 [-1.91, 0.51] |

**Comparison 4. ORT-based algorithm vs standard dose or non-ORT based algorithm**

| Outcome or subgroup title | No. of studies | No. of participants | Statistical method | Effect size |
|---|---|---|---|---|
| 1 Live birth or ongoing pregnancy | 4 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 1.1 Control group: standard dose 150 IU | 4 | 2823 | Odds Ratio (M-H, Fixed, 95% CI) | 1.04 [0.88, 1.23] |
| 2 Severe OHSS | 4 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 2.1 Control group: standard dose 150 IU | 3 | 1494 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.54 [0.14, 1.99] |
| 2.2 Control group: non-ORT algorithm | 1 | 194 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 3 Clinical pregnancy | 5 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |

| | | | | |
|---|---|---|---|---|
| 3.1 Control group: standard dose 150 IU | 4 | 2823 | Odds Ratio (M-H, Fixed, 95% CI) | 0.96 [0.82, 1.13] |
| 3.2 Control group: non-ORT algorithm | 1 | 194 | Odds Ratio (M-H, Fixed, 95% CI) | 0.88 [0.48, 1.61] |
| 4 Time to clinical pregnancy | 0 | 0 | Mean Difference (IV, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 5 Moderate or severe OHSS | 5 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 5.1 Control group: standard dose 150 IU | 4 | 2823 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.58 [0.34, 1.00] |
| 5.2 Control group: non-ORT algorithm | 1 | 194 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 6 Multiple pregnancy in randomised women | 4 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 6.1 Control group: standard dose 150 IU | 4 | 2823 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.77 [0.43, 1.36] |
| 6.2 Control group: non-ORT algorithm | 0 | 0 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 7 Log(N oocytes retrieved); data presented on the log scale - cannot interpret pooled estimates as N of oocytes | 5 | | Mean Difference (Fixed, 95% CI) | Subtotals only |
| 7.1 Control group: standard dose 150 IU | 4 | | Mean Difference (Fixed, 95% CI) | 0.16 [0.11, 0.20] |
| 7.2 Control group: non-ORT algorithm | 1 | | Mean Difference (Fixed, 95% CI) | 0.12 [-0.02, 0.26] |
| 8 Poor response to stimulation | 3 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 8.1 Control group: standard dose 150 IU | 2 | 1294 | Odds Ratio (M-H, Fixed, 95% CI) | 1.16 [0.90, 1.50] |
| 8.2 Control group: non-ORT algorithm | 1 | 194 | Odds Ratio (M-H, Fixed, 95% CI) | 0.50 [0.27, 0.92] |
| 9 Normal response to stimulation | 4 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 9.1 Control group: standard dose 150 IU | 3 | 2623 | Odds Ratio (M-H, Fixed, 95% CI) | 1.22 [1.04, 1.43] |
| 9.2 Control group: non-ORT algorithm | 1 | 194 | Odds Ratio (M-H, Fixed, 95% CI) | 2.13 [1.20, 3.78] |
| 10 Hyper-response to stimulation | 3 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 10.1 Control group: standard dose 150 IU | 2 | 1294 | Odds Ratio (M-H, Fixed, 95% CI) | 0.56 [0.42, 0.76] |
| 10.2 Control group: non-ORT algorithm | 1 | 194 | Odds Ratio (M-H, Fixed, 95% CI) | 0.57 [0.25, 1.31] |
| 11 Cycle cancellations for poor response | 5 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 11.1 Control group: standard dose 150 IU | 4 | 2823 | Odds Ratio (M-H, Fixed, 95% CI) | 1.19 [0.89, 1.60] |
| 11.2 Control group: non-ORT algorithm | 1 | 194 | Odds Ratio (M-H, Fixed, 95% CI) | 0.14 [0.01, 2.83] |
| 12 Cycle cancellations for hyper-response | 5 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 12.1 Control group: standard dose 150 IU | 4 | 2823 | Odds Ratio (M-H, Fixed, 95% CI) | 0.37 [0.24, 0.57] |
| 12.2 Control group: non-ORT algorithm | 1 | 194 | Odds Ratio (M-H, Fixed, 95% CI) | 0.95 [0.40, 2.27] |

| | | | | |
|---|---|---|---|---|
| 13 Cycle cancellations for poor or hyper-response | 5 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 13.1 Control group: standard dose 150 IU | 4 | 2823 | Odds Ratio (M-H, Fixed, 95% CI) | 0.78 [0.61, 1.00] |
| 13.2 Control group: non-ORT algorithm | 1 | 194 | Odds Ratio (M-H, Fixed, 95% CI) | 0.73 [0.32, 1.69] |
| 14 Women with at least one transferable embryo | 5 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 14.1 Control group: standard dose 150 IU | 4 | 2823 | Odds Ratio (M-H, Fixed, 95% CI) | 0.90 [0.74, 1.10] |
| 14.2 Control group: non-ORT algorithm | 1 | 194 | Odds Ratio (M-H, Fixed, 95% CI) | 1.15 [0.57, 2.33] |
| 15 Total dose of FSH | 4 | | Mean Difference (IV, Fixed, 95% CI) | Subtotals only |
| 15.1 Control group: standard dose 150 IU | 3 | 1494 | Mean Difference (IV, Fixed, 95% CI) | -155.00 [-215.54, -98.45] |
| 15.2 Control group: non-ORT algorithm | 1 | 194 | Mean Difference (IV, Fixed, 95% CI) | -11.0 [-210.30, 188.30] |
| 16 Duration of FSH administration | 5 | | Mean Difference (IV, Fixed, 95% CI) | Subtotals only |
| 16.1 Control group: standard dose 150 IU | 4 | 2823 | Mean Difference (IV, Fixed, 95% CI) | 0.23 [0.09, 0.37] |
| 16.2 Control group: non-ORT algorithm | 1 | 194 | Mean Difference (IV, Fixed, 95% CI) | -0.40 [-0.84, 0.04] |
| 17 Cost per woman randomised | 0 | 0 | Mean Difference (IV, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 18 Cumulative live birth: 1 cycle (fresh + frozen) | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 18.1 Control group: standard dose 150 IU | 1 | 1032 | Odds Ratio (M-H, Fixed, 95% CI) | 0.87 [0.66, 1.14] |
| 18.2 Control group: non-ORT algorithm | 0 | 0 | Odds Ratio (M-H, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 19 Cumulative live birth: 18 months | 1 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 19.1 Control group: standard dose 150 IU | 1 | 1032 | Odds Ratio (M-H, Fixed, 95% CI) | 0.89 [0.70, 1.14] |
| 19.2 Control group: non-ORT algorithm | 0 | 0 | Odds Ratio (M-H, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 20 Multiple pregnancy in women with clinical pregnancy | 4 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 20.1 Control group: standard dose 150 IU | 4 | 898 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.71 [0.39, 1.28] |
| 20.2 Control group: non-ORT algorithm | 0 | 0 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |

**Comparison 5. AMH-based algorithm vs AFC-based algorithm**

| Outcome or subgroup title | No. of studies | No. of participants | Statistical method | Effect size |
|---|---|---|---|---|
| 1 Live birth or ongoing pregnancy | 0 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 2 Severe OHSS | 1 | 348 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 3 Clinical pregnancy | 1 | 348 | Odds Ratio (M-H, Fixed, 95% CI) | 0.82 [0.53, 1.27] |
| 4 Time to clinical pregnancy | 0 | 0 | Mean Difference (IV, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 5 Moderate or severe OHSS | 1 | 348 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 4.28 [0.96, 19.07] |
| 6 Multiple pregnancy in randomised women | 1 | 348 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 1.21 [0.66, 2.23] |
| 7 Log(N oocytes retrieved); data presented on the log scale - cannot interpret pooled estimates as N of oocytes | 1 | | Mean Difference (Fixed, 95% CI) | -0.25 [-0.37, -0.13] |
| 8 Poor response to stimulation | 1 | | Odds Ratio (Fixed, 95% CI) | 2.25 [0.94, 5.35] |
| 9 Normal response to stimulation | 1 | | Odds Ratio (Fixed, 95% CI) | 1.38 [0.87, 2.17] |
| 10 Hyper-response to stimulation | 1 | | Odds Ratio (Fixed, 95% CI) | 0.45 [0.23, 0.88] |
| 11 Cycle cancellations for poor response | 1 | 348 | Odds Ratio (M-H, Fixed, 95% CI) | 1.51 [0.25, 9.14] |
| 12 Cycle cancellations for hyper-response | 1 | 348 | Odds Ratio (M-H, Fixed, 95% CI) | 0.54 [0.23, 1.25] |
| 13 Cycle cancellations for poor or hyper-response | 1 | 348 | Odds Ratio (M-H, Fixed, 95% CI) | 0.64 [0.30, 1.38] |
| 14 Women with at least one transferable embryo | 1 | 348 | Odds Ratio (M-H, Fixed, 95% CI) | 1.21 [0.36, 4.03] |
| 15 Total dose of FSH | 1 | 348 | Mean Difference (IV, Fixed, 95% CI) | -178.0 [-413.88, 57.88] |
| 16 Duration of FSH administration | 1 | 348 | Mean Difference (IV, Fixed, 95% CI) | 0.20 [-0.11, 0.51] |
| 17 Live birth or ongoing pregnancy (including FET for freeze-all) | 0 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 18 Cost per woman randomised | 0 | 0 | Mean Difference (IV, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 19 Multiple pregnancy in women with clinical pregnancy | 1 | 128 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 1.59 [0.78, 3.25] |

**Comparison 6. AMH + AFC-based algorithm vs AFC-based algorithm**

| Outcome or subgroup title | No. of studies | No. of participants | Statistical method | Effect size |
|---|---|---|---|---|
| 1 Live birth or ongoing pregnancy | 1 | 308 | Odds Ratio (M-H, Fixed, 95% CI) | 1.18 [0.72, 1.93] |
| 2 Severe OHSS | 1 | 308 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.68 [0.12, 4.00] |
| 3 Clinical pregnancy | 1 | 308 | Odds Ratio (M-H, Fixed, 95% CI) | 1.37 [0.84, 2.23] |
| 4 Time to clinical pregnancy | 0 | 0 | Mean Difference (IV, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 5 Moderate or severe OHSS | 1 | 308 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.85 [0.26, 2.83] |

| | | | | |
|---|---|---|---|---|
| 6 Multiple pregnancy in randomised women | 1 | 308 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 7 Log(N oocytes retrieved); data presented on the log scale - cannot interpret pooled estimates as N of oocytes | 1 | | Mean Difference (Fixed, 95% CI) | -0.19 [-0.31, -0.07] |
| 8 Poor response to stimulation | 1 | | Odds Ratio (Fixed, 95% CI) | 2.82 [1.52, 5.25] |
| 9 Normal response to stimulation | 1 | | Odds Ratio (Fixed, 95% CI) | 0.71 [0.45, 1.12] |
| 10 Hyper-response to stimulation | 1 | | Odds Ratio (Fixed, 95% CI) | 0.72 [0.43, 1.23] |
| 11 Cycle cancellations for poor response | 1 | 308 | Odds Ratio (M-H, Fixed, 95% CI) | 1.83 [0.53, 6.40] |
| 12 Cycle cancellations for hyper-response | 1 | 308 | Odds Ratio (M-H, Fixed, 95% CI) | 0.50 [0.12, 2.05] |
| 13 Cycle cancellations for poor or hyper-response | 1 | 308 | Odds Ratio (M-H, Fixed, 95% CI) | 1.03 [0.42, 2.55] |
| 14 Women with at least one transferable embryo | 1 | 308 | Odds Ratio (M-H, Fixed, 95% CI) | 1.04 [0.50, 2.19] |
| 15 Total dose of FSH | 1 | 308 | Mean Difference (IV, Fixed, 95% CI) | 81.0 [-111.93, 273.93] |
| 16 Duration of FSH administration | 1 | 308 | Mean Difference (IV, Fixed, 95% CI) | 0.5 [0.10, 0.90] |
| 17 Cost per woman randomised | 0 | 0 | Mean Difference (IV, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 18 Live birth or ongoing pregnancy (including FET for freeze-all) | 1 | 308 | Odds Ratio (M-H, Fixed, 95% CI) | 1.25 [0.77, 2.05] |
| 19 Multiple pregnancy in women with clinical pregnancy | 1 | 83 | Peto Odds Ratio (Peto, Fixed, 95% CI) | 0.0 [0.0, 0.0] |

**Comparison 7. AMH + AFC + bFSH-based algorithm vs bFSH-based algorithm**

| Outcome or subgroup title | No. of studies | No. of participants | Statistical method | Effect size |
|---|---|---|---|---|
| 1 Live birth or ongoing pregnancy | 1 | 215 | Odds Ratio (M-H, Fixed, 95% CI) | 0.54 [0.28, 1.04] |
| 2 Severe OHSS | 0 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 3 Clinical pregnancy | 1 | 215 | Odds Ratio (M-H, Fixed, 95% CI) | 0.51 [0.28, 0.93] |
| 4 Time to clinical pregnancy | 0 | 0 | Mean Difference (IV, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 5 Moderate or severe OHSS | 0 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 6 Multiple pregnancy in randomised women | 0 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |
| 7 Log(N oocytes retrieved); data presented on the log scale - cannot interpret pooled estimates as N of oocytes | 1 | | Mean Difference (Fixed, 95% CI) | -0.20 [-0.81, 0.41] |
| 8 Poor response to stimulation | 1 | | Odds Ratio (Fixed, 95% CI) | 1.46 [0.77, 2.79] |
| 9 Normal response to stimulation | 1 | | Odds Ratio (Fixed, 95% CI) | 0.75 [0.42, 1.35] |
| 10 Hyper-response to stimulation | 1 | | Odds Ratio (Fixed, 95% CI) | 0.93 [0.45, 1.93] |
| 11 Cycle cancellations for poor response | 0 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |

| | | | | |
|---|---|---|---|---|
| 12 Cycle cancellations for hyper-response | 0 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 13 Cycle cancellations for poor or hyper-response | 0 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 14 Women with at least one transferable embryo | 0 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 15 Total dose of FSH | 1 | 215 | Mean Difference (IV, Fixed, 95% CI) | -148.0 [-433.61, 137.61] |
| 16 Duration of FSH administration | 1 | 215 | Mean Difference (IV, Fixed, 95% CI) | 0.0 [-0.60, 0.60] |
| 17 Cost per woman randomised | 0 | 0 | Mean Difference (IV, Fixed, 95% CI) | 0.0 [0.0, 0.0] |
| 18 Live birth or ongoing pregnancy (including FET for freeze-all) | 0 | | Odds Ratio (M-H, Fixed, 95% CI) | Subtotals only |
| 19 Multiple pregnancy in women with clinical pregnancy | 0 | | Peto Odds Ratio (Peto, Fixed, 95% CI) | Subtotals only |

## CONTRIBUTIONS OF AUTHORS

SL and JW developed the protocol and search strategies, with input and final approval from all protocol authors (JL joined the team at the review stage). SL, JW and JL performed the screening, data extraction, risk of bias assessments, and meta-analysis. SL and JW drafted the review. All authors read and commented on the draft versions of the review and approved the final version.

## DECLARATIONS OF INTEREST

Sarah Lensen does not have any conflicts of interest to declare.

Jack Wilkinson was funded by a Doctoral Research Fellowship supported by the National Institute for Health Research while completing part of this work (DRF-2014-07-050). The views expressed in this publication are those of the authors and not necessarily those of the National Health Service (NHS), the National Institute for Health Research or the Department of Health. JW also declares that publishing research is beneficial to his career.

Jori Leijdekkers received an unrestricted personal grant from Merck B.V.

Antonio La Marca received research funding to his institution from companies producing gonadotrophins (Merck Serono, Ferring, IBSA, MSD). He did not receive direct payment for consultancies or conference speaking in the last 3 years from any of these companies. Antonio La Marca is an investigator on a trial included in this review (Allegra 2017). To manage this potential conflict, review authors did not extract data or assess risk of bias for the studies they have been involved with.

Ben Mol has received consultancy fees from ObsEva, Guerbet and Merck; payment for review preparation from European Journal of Obstetrics and Gynecology and Reproductive Biology; and travel/accommodation/meeting expenses for various non-commercial scientific meetings.

Jane Marjoribanks does not have any conflicts of interest to declare.

Helen Torrance has received travel grants from from Merck B.V.

Frank Broekmans has not consulted for any related companies in the last 3 years.

Frank Broekmans, Ben Mol and Helen Torrance are investigators in the OPTIMIST trial, which is included in this review (Oudshoorn 2017; Van Tilborg 2017). Frank Broekmans is an investigator on two other trials included in this review (Klinkert 2005; Olivennes 2015). To manage this potential conflict, review authors did not extract data or assess risk of bias for the studies they have been involved with.

## SOURCES OF SUPPORT

### Internal sources

- None, Other.

### External sources

- NIHR Cochrane Incentive Award, UK.

The award is made available at submission of the completed review within the agreed timeframe.

## DIFFERENCES BETWEEN PROTOCOL AND REVIEW

There are a number of differences between the review and protocol, most of which we describe in Potential biases in the review process. The largest changes were to the structure of the review and the use of meta-analysis. Although we did not intend to pool direct dose comparison studies using different dose sets, we made a post hoc decision to pool some of these. Further, we shifted one of the included trials, Van Tilborg 2017 from the poor responder comparison to the normal responder comparison after discussion with the review authors.

We included an outcome of multiple pregnancy in women with clinical pregnancy after the protocol stage, which we acknowledge is associated with collider bias, but which we considered to provide a more useful rate of the outcome.

We re-ordered the outcomes (shifting the outcome of 'moderate or severe OHSS' up).