Cochrane Gynaecology and Fertility

COCHRANE GYNAECOLOGY AND FERTILITY GROUP GUIDANCE

This guidance is in four sections, in the order that they appear in a published review.

- For a new protocol, read part A below then go to part D
- For a full review or update, use all sections
- Examples below are in italics, while notes and comments from us are in plain text.

Index

PART A: ALL AUTHORS PLEASE READ - IT WILL SAVE YOU TIME!

By far the most common problems with reviews involve the issues listed below:

Methods

- It is mandatory for review authors to search the registers of ongoing trials, such as http://www.clinicaltrials.gov). The Trials Search Coordinator (Information Specialist) does not do this.
- In general, do not exclude studies on the basis of their reported outcome measures. Include studies that *could* have measured the outcomes of interest even if they did not.
- Define in advance details of what are acceptable outcome measures (e.g. differing scales, time-points), and state a preference order when there are several possible measures.
- Keep subgroup analyses to an absolute minimum, and explain the rationale for these.
- If planning any departures from Methods specified in protocol, check with editorial base first; document and justify any changes in the Changes from Protocol section.

Reporting results

- Include all pre-specified comparisons and outcomes: if there are no relevant data, say so
- Use this format for presenting results: (RR 0.89, 95% CI 0.75 to 1.05, three RCTs, n=811, I^2=31%, low quality evidence).
- Do not confuse lack of evidence of an effect with evidence of a lack of effect: say "There was no evidence of a difference' not "There was no significant difference"
- Consider clinical rather than statistical significance
- Interpret subgroups very cautiously. In general, do not report them in the abstract.
- If investigating subgroup differences, present an overall plot or figure containing all subgroups, rather than multiple forest plots.
- Include the sample size for each included study and for each intervention group in the Characteristics of Included Studies table

Conclusions

In summary parts of the review Abstract, Plain language summary and SoF:

- include the same outcomes: include all primary outcomes and adverse events
- include the same comparisons: those that are clinically most important. These should be specified at the protocol stage, rather than on the basis of the results.
- incorporate the findings of the GRADE assessment.
- be 100% consistent

Summary of findings table

- Specify detailed plans for the SoF table in the Methods section (see page 13 below).
- The SoF should be prepared once the study data have been entered, and before the results section is written
- Clearly explain Summary of Findings (SoF) evidence downgrades in footnotes: e.g. We downgraded the evidence by two levels, due to very serious imprecision: only 29 events
- When you reach this stage, before proceeding any further please submit your review to our Managing Editor for an editorial check
- For help with preparing a summary of findings table, contact us or see http://tech.cochrane.org/revman/other-resources/gradepro/resources.
  Contact details for Managing Editor: h.nagels@auckland.ac.nz

PART B - Abstract

Background

Treatment B is commonly used to optimize the chance of live birth in women undergoing assisted reproductive technology (ART). However, it is known to increase multiple pregnancy rates, potentially causing maternal and perinatal morbidity. Treatment A is an alternative intervention which may reduce the risk of multiple pregnancy. We compared the benefits and risks of the two treatments

Objectives

To evaluate the effectiveness and safety of Treatment A in women undergoing ART.

Search Methods

The Cochrane Gynaecology and Fertility (CGF) Group trials register, CENTRAL, MEDLINE, Embase, PsycINFO, CINAHL and two trials registers were searched in August 2016 together with reference checking and contact with study authors and experts in the field to identify additional studies.

Selection criteria

We included randomized controlled trials (RCTs) of the use of Treatment A compared with Treatment B for subfertile women.

Data collection and analysis

We used standard methodological procedures recommended by The Cochrane Collaboration. The primary review outcomes were cumulative live birth and multiple pregnancy. Other adverse effects were a secondary outcome.

Main results

We included three RCTs (811 women analysed) comparing Treatment A with Treatment B. The evidence was low to moderate quality: the main limitations were serious risk of bias due to poor reporting of study methods, and serious imprecision.

Live birth

There was no conclusive evidence of a difference in cumulative live birth rate between Treatment A and Treatment B (RR 0.89, 95% CI 0.75 to 1.05, three RCTs, n=811, I2=0%, low quality evidence). This suggests that if the chance of live birth following Treatment B is assumed to be 42%, the chance following Treatment A would be between 31% and 44%.

Multiple pregnancy

Treatment A was associated with lower multiple pregnancy rates than Treatment B (RR 0.04, 95% CI 0.01 to 0.15, three RCTs, n = 811, I2 = 23%, moderate quality evidence). This suggests that if the risk of multiple pregnancy following Treatment B is assumed to be 13%, the risk following Treatment A would be between 0% and 2%.

Other adverse effects

There was insufficient evidence to reach a conclusion regarding other adverse effects, as no studies reported data suitable for analysis.

Authors' conclusions

Evidence suggests that Treatment A may minimise the risk of multiple pregnancy in women undergoing ART without substantially reducing the cumulative live birth rate. Data are lacking on other adverse effects.

- If reporting any secondary outcomes in the Abstract, choose them by clinical importance - do not "cherry pick" statistically significant findings.
- Do not report subgroup analyses in the abstract
- Aim to limit the abstract to fewer than 700 words. Absolute maximum is 1000

PART C - Plain language summary

Title: Treatment A versus treatment B for women undergoing assisted reproductive technology (ART)

Review question

Researchers in the Cochrane Collaboration reviewed the evidence about the effect of Treatment A versus Treatment B in women undergoing ART.

Background

Treatment B is commonly used to increase the chance of live birth in women undergoing ART. However, it is known to increase multiple pregnancy rates, which can cause serious health risks for both mother and baby. Treatment A is an alternative approach which may reduce the risk of multiple pregnancy. We compared the benefits and risks of the two treatments.

Study characteristics

We found 3 randomised controlled trials comparing Treatment A with Treatment B in a total of 811 women undergoing ART. The evidence is current to June 2016.

Key results

Treatment A may minimise the risk of multiple pregnancy in women undergoing ART without substantially reducing the cumulative live birth rate. The evidence suggests that if the chance of live birth following Treatment B is assumed to be 42%, the chance following Treatment A would be between 31% and 44%). It also suggests that if the risk of multiple pregnancy following Treatment B is assumed to be 13%, the risk following Treatment A would be between 0% and 2%.

Evidence on other adverse events was poorly reported and inconclusive.

Quality of the evidence

The evidence was of low to moderate quality. The main limitations in the evidence were poor reporting of study methods, and lack of precision in the findings for live birth.

- Format the plain language summary under the five headings used above
- For more detailed information on the standards for Plain language summaries go to: *Standards for the reporting of Plain Language Summaries in new Cochrane Intervention Reviews*
- Translate the effect estimates for important clinical outcomes into language that uses natural frequencies. Rates per 100 (as used above) are easily extracted from the review's Summary of Findings Table.
- Report and interpret the units used for continuous outcomes (e.g. a VAS scale of 0-10 where 0 is pain-free and 10 is unbearable pain)
- Report all primary outcomes in the Abstract and PLS, and summarise any evidence about adverse effects (including lack of data).
- Aim to limit the abstract to fewer than 700 words. Absolute maximum is 1000

PART D– Methods

Criteria for considering studies for this review

Types of studies

Randomised controlled trials (RCTs) will be included

- Usually all randomised trial designs should be included. A common exception is crossover trials in situations where treatment success precludes "crossover" (e.g. in subfertility trials, where the goal is pregnancy/live birth). Various acceptable options exist:
- Crossover trials will be excluded, as the design is not valid in this context
- Crossover trials will be included but only data from the first phase will be included in meta-analyses, as the crossover is not a valid design in this context.
- Crossover trials will be included, as the crossover is a valid design in this context
- Define any potentially ambiguous terms, such as "double-blind".

Types of participants

Women/couples with unexplained infertility will be eligible for inclusion.

- Definitions of the condition (e.g. unexplained infertility) belong in the Background section.

Types of interventions

Trials comparing Factor X via any route versus any other active intervention or placebo will be included

Types of outcome measures

Primary outcomes

1. Live birth or ongoing pregnancy
    - *Live birth is defined as delivery of a live fetus after 20 completed weeks of gestation*
    - *Ongoing pregnancy is defined as evidence of a gestational sac with fetal heart motion at 12 weeks, confirmed with ultrasound*
2. *Multiple birth*

Secondary outcomes

3. *Clinical pregnancy, defined as evidence of a gestational sac, confirmed by ultrasound.*
4. *Any adverse event (including miscarriage, bleeding, drug reactions), reported either as a composite measure or separately.*
5. *Quality of life. If studies report more than one scale, preference will be given to the SF-36, then other validated generic scales, and finally condition-specific scales.*

- In general, avoid excluding studies based on outcomes. Include eligible studies that could have measured the outcomes of interest even if they did not.
- The primary outcomes should be as few as possible and should normally include one measure of effectiveness and one of potential harm. The review conclusions will be based primarily on these outcomes.
- Consider which seven outcomes will be included in the summary of findings table.
-

- Number outcomes as in the example above, and use the same numbers in the Effects of Interventions section.
- Keep secondary outcomes to a minimum. Focus on clinical outcomes and try to avoid lab outcomes (e.g. implantation rate).
- Define in advance details of what are acceptable outcome measures (e.g. differing definitions, assessors, scales, time-points) and state a preference order when there are several possible measures.

Search methods for identification of studies

We will search for all published and unpublished RCTs of XX, without language restriction and in consultation with the Gynaecology and Fertility Group (CGF) Information Specialist:

Electronic searches

(1) We will search the following electronic databases for relevant trials:

2. *The Cochrane Gynaecology and Fertility Group (CGF) Specialised Register of Controlled Trials, PROCITE platform (from inception onwards)*
3. *The Cochrane Central Register of Controlled Trials; via the Cochrane Register of Studies Online (CRSO Web platform) (from inception onwards)*

MEDLINE Epub Ahead of Print, In-Process & Other Non-Indexed Citations) Ovid (from 1946 onwards);

4. *Embase Ovid (1974 onwards);*
5. *PsycINFO Ovid (from 1806 onwards)*
6. *CINAHL; (Cumulative Index to Nursing and Allied Health Literature; (from 1982 onwards)*

The MEDLINE search will be combined with the Cochrane highly sensitive search strategy for identifying randomized trials which appears in the Cochrane Handbook of Systematic Reviews of Interventions (Version 5.1.0 chapter 6, 6.4.11). The Embase, PsycINFO and CINAHL searches are combined with trial filters developed by the Scottish Intercollegiate Guidelines Network (SIGN) http://www.sign.ac.uk/methodology/filters.html#random.

(2) *Other electronic sources of trials will include*

7. *trial registers for ongoing and registered trials–*
    - http://www.clinicaltrials.gov *(a service of the US National Institutes of Health)*
    - http://www.who.int/trialsearch/Default.aspx *(The World Health Organisation International Trials Registry Platform search portal)*
      Note: it is now mandatory for Cochrane reviews to include searches of trial registers
8. *LILACS and other Spanish,Portuguese language databases (Latin American and Caribbean Health Science Information database (from 1982 ongoing). found in the Virtual Health Library Regional Portal (VHL)* http://bvsalud.org/portal/?lang=en *(the right hand drop down box allows you to filter out MEDLINE records). .*

- *PubMed and Google Scholar (for recent trials not yet indexed in the major databases)*

[The searches above will be simple short keyword searches and should also be documented in the appendices]

Searching other resources

(3) *We will hand search reference lists of relevant trials and systematic reviews retrieved by the search and contact experts in the field to obtain additional data. We will also hand search relevant journals and conference abstracts that are not covered in the CGF register, in liaison with the Information Specialist.*

Gynaecology and Fertility search requirements:

1. Designing and running the search

- The Gynaecology and Fertility Information Specialist (Marian Showell) will help design your search and will run a search in the electronic databases listed under (1) above. It is the responsibility of the review authors to run, document and date (with day, month and year) the searches of other sources (i.e. those listed under (2) and (3) above, as appropriate).

2. Sources that all review authors MUST search:

- Trial registries (a very important source for recent and ongoing trials)
- Reference lists of articles retrieved
- Reviews of Traditional Chinese Medicine or Chinese complementary therapies must search at least one Chinese database

3. Resources for advice on searching

- Liaise with the CG F Information Specialist to avoid duplication of handsearching and for other advice on searching
- Guidance on search strategies, accessing specific databases (including LILACS and Chinese databases), and lists of journals can also be found on the CGF Module at: http://onlinelibrary.wiley.com/o/cochrane/clabout/articles/MENSTR/sect0-meta.html.

4. Documenting the search

- List all sources searched in the Methods section of the review (as in the example above).
- The search process should be summarised in a PRISMA flow diagram in the full review.
- Full search strategies for all sources searched must be copied and pasted into the appendices (not in the body of the text) of the review along with dates and the platforms used for each database.
- The MEDLINE, Embase and PsycINFO searches are on the OVID platform, CENTRAL is now searched via CRS ONLINE via the Web, while CINAHL is currently searched on the EBSCO platform.

In the protocol the numbers of hits per search line (i.e. the numbers in brackets after the keywords) are removed from the strategies, however at the review stage the numbers of hits per keyword remain.

5. *Ensuring the search is current.*

- The 'assessed as up to date' date in the header of your review must be the same as the date of the searches that have been incorporated in the review.
- It is mandatory to run/update searches for all relevant databases no more than six months (maximum 12 months) before publication of the full review.
- Ideally any new studies should be fully incorporated. As a minimum, potentially eligible studies should be referenced under "awaiting classification" or "ongoing", but authors have to show that they are waiting for information from trial authors.

Data collection and analysis

There should not be any text under this heading.

Selection of studies

After an initial screen of titles and abstracts retrieved by the search, conducted by XXX, we will retrieve the full texts of all potentially eligible studies. Two review authors (XX and YY) will independently examine these full text articles for compliance with the inclusion criteria and select eligible studies. We will correspond with study investigators as required, to clarify study eligibility. Disagreements will be resolved by discussion. If any reports require translation, we will describe the process used for data collection. We will document the selection process with a "PRISMA" flow chart.*

- *It is preferable (but not mandatory) that two people independently do this initial screen

Data extraction and management

Two review authors will independently extract data from eligible studies using a data extraction form designed and pilot-tested by the authors. Any disagreements will be resolved by discussion. Data extracted will include study characteristics and outcome data (see data extraction table for details, Appendix XX). Where studies have multiple publications the authors will collate multiple reports of the same under a single study ID with multiple references.

We will correspond with study investigators for further data on methods and/or results, as required.

- Data are often presented in a non-standardised format. Studies should be included irrespective of whether outcomes are reported in a "usable" way. In multi-arm studies, data from arms that do not meet eligibility criteria should be excluded.

Assessment of risk of bias in included studies

Two review authors will independently assess the included studies for risk of bias using the Cochrane risk of bias assessment tool (Higgins 2011)) to assess: 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 bias. Judgements will be assigned as recommended in the Cochrane Handbook Section 8.5 (Higgins 2011). Disagreements will be resolved by discussion. We will describe all judgements fully and present the conclusions in the Risk of Bias table, which will be incorporated into the interpretation of review findings by means of sensitivity analyses (see below)

With respect to within-trial selective reporting, where identified studies fail to report the primary outcome of live birth, but do report interim outcomes such as pregnancy, we will assess whether the interim values are similar to those reported in studies that also report live birth.

- Read section 8.5 of the handbook for detailed guidance on assessing each type of bias
- If likely sources of "other bias" can be identified in advance, these should be specified in this section and the number of domains increased accordingly.
- Assessment of risk of bias involves considering the potential impact of each domain in the context of individual studies (or even individual outcomes). For example, lack of blinding

may not increase the risk of bias if follow-up is complete and outcomes are unequivocal (e.g. live birth).

- Selective reporting is a type of reporting bias that affects the internal validity of an individual study. It refers to the selective reporting of some outcomes (e.g. positive outcomes) and the failure to report others (e.g. adverse events). Trialists should report all pre-stated outcomes, which should include all outcomes that you would expect, such as adverse events.

Measures of treatment effect

For dichotomous data (e.g. live birth rates), we will use the numbers of events in the control and intervention groups of each study to calculate Mantel-Haenszel odds ratios (ORs). For continuous data (e.g. weight gain), if all studies report exactly the same outcomes we will calculate mean difference (MDs) between treatment groups. If similar outcomes are reported on different scales (e.g. change in weight) we will calculate the standardised mean difference (SMD). We will reverse the direction of effect of individual studies, if required, to ensure consistency across trials. We will treat ordinal data (e.g. quality of life scores) as continuous data. We will present 95% confidence intervals for all outcomes. Where data to calculate ORs or MDs are not available, we will utilise the most detailed numerical data available that may facilitate similar analyses of included studies (e.g. test statistics, p values). We will assess whether the estimates calculated in the review for individual studies are compatible in each case with the estimates reported in the study publications.*

- *There are three available statistics to analyse binary (dichotomous) outcomes - the odds ratio, risk ratio and risk difference. The odds ratio further divides into the Mantel-Haenszel and Peto estimates. Any analysis compatible with the Handbook (please see chapter 9) is acceptable although it is rarely appropriate to use a risk difference. We recommend use of the odds ratio (Mantel Haenszel by default, Peto if events are very rare) because of its superior mathematical properties. Whichever statistic you use, we encourage 'translation' of the result to actual percentages for a typical population to maximise understanding. You will find examples of this in Sections B, C and E.
- Only include information relevant to the review (e.g. many subfertility reviews contain only binary outcomes, so you do not need to provide for continuous outcomes.)

Unit of analysis issues

*The primary analysis will be per woman randomised; per pregnancy data may also be included for some outcomes (e.g. miscarriage). Data that do not allow valid analysis (e.g. "per cycle" data) will be briefly summarised in an additional table and will not be meta-analysed. Multiple births will be counted as one live birth event. Only first-phase data from crossover trials will be included. [*OR*: Statistical advice will be sought regarding the analysis of crossover trials, to facilitate the appropriate inclusion of crossover data in meta-analysis].*

- If studies report only "per cycle" data, contact authors and request "per woman" data.
- Some outcomes can only occur in women who reach clinical pregnancy (e.g. multiple pregnancy, miscarriage, etc) Report all outcomes per randomised woman, as this is the unit of randomisation. Rates per clinical pregnancy may be used as the denominator for a secondary analysis, as this will help give the full picture.

Dealing with missing data

We will analyse the data on an intention-to-treat basis as far as possible (i.e. including all randomised participants in analysis, in the groups to which they were randomised). Attempts will be made to obtain missing data from the original trialists. Where these are unobtainable, we will undertake imputation of individual values for live birth only: live birth will be assumed not to have occurred in participants without a reported outcome. For other outcomes, we will analyse only the available data. Any imputation undertaken will be subjected to sensitivity analysis (see below).

If studies report sufficient detail to calculate mean differences but no information on associated standard deviation (SD), we will assume the outcome to have a standard deviation equal to the highest SD from other studies within the same analysis.

Assessment of heterogeneity

We will consider whether the clinical and methodological characteristics of the included studies are sufficiently similar for meta-analysis to provide a clinically meaningful summary. We will assess statistical heterogeneity by the measure of the I2. An I^2 measurement greater than 50% will be taken to indicate substantial heterogeneity (Higgins 2011)

Assessment of reporting biases

In view of the difficulty of detecting and correcting for publication bias and other reporting biases, we will aim to minimise their potential impact by ensuring a comprehensive search for eligible studies and by being alert for duplication of data. If there are ten or more studies in an analysis, we will use 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).

- This section refers to review-wide reporting bias (e.g. publication bias, multiple publication bias, language bias etc), whereby the dissemination of research findings is influenced by the nature and direction of results, reducing the likelihood that all studies eligible for a review will be retrieved.

Data synthesis

If the studies are sufficiently similar, we will combine the data using a fixed effect/random effects model in the following comparisons:*

1. *Factor X versus placebo, subgrouped by mode of administration and dose. We plan to pool the data*

(i) *Low dose oral*

(ii) *High dose oral*

2. *Factor X versus Factor G*
3. *Factor X versus Factor H*

- Define analyses that are comprehensive and mutually exclusive, so that all results can be slotted into one category only, and so that trials within the same category can sensibly be pooled. This allows consideration of effects within each category as well as, or instead of, an overall estimate for the comparison.
- If analyses are subgrouped (as in example 1 above), state whether it is planned to pool the subgroups
- *In general, fixed-effect analysis can be used if it is reasonable to assume that the underlying effect size is the same for all the trials in the analysis. Otherwise consider using random-effects analysis to obtain an overall summary, or do not combine the trials. It is important that you can justify whatever decision you make.

Statistical analysis will be performed using Review Manager 5.3 (RevMan 2014).

Subgroup analysis and investigation of heterogeneity

Where data are available, we will conduct subgroup analyses to determine the separate evidence within the following subgroups:

1. *Studies of low dose versus studies of high dose*
2. *Studies including only women with a high BMI (>32kg/m2)*

If we detect substantial heterogeneity, we will explore possible explanations in subgroup analyses (e.g. differing populations) and/or sensitivity analyses (e.g. differing risk of bias). We will take any statistical heterogeneity into account when interpreting the results, especially if there is any variation in the direction of effect.

- Keep subgroups to an absolute minimum
- Subgroups can be defined either by characteristics of the study or by those of the participants. In practice, the latter are unlikely to be available in reported data. Subgroups should be explicit and few. Preferably the rationale for each will be clear from the Background section. If not, it must be explained here.
- If subgroups are to be compared, this should be done with a formal statistical test. Interpretation of the statistical analysis for subgroups is problematic.

Sensitivity analysis

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

1. *Eligibility had been restricted to studies without high risk of bias*
2. *A [fixed effect/random effects (delete one!)] model had been adopted*
3. *Alternative imputation strategies had been implemented*
4. *The summary effect measure had been relative risk rather than odds ratio*

Overall quality of the body of evidence: Summary of findings table

We will prepare a Summary of findings table using GRADEpro and Cochrane methods. This table will evaluate the overall quality of the body of evidence for the main review outcomes (Live birth, ongoing pregnancy, multiple pregnancy, OHSS) for the main review comparison (Treatment A versus placebo). Additional Summary of Findings tables will be also prepared for the main review outcomes for other important comparisons (Treatment A versus Treatment B, and Treatment B versus Treatment C). We will assess the quality of the evidence using GRADE criteria: risk of bias, consistency of effect, imprecision, indirectness and publication bias). Judgements about evidence quality (high, moderate, low or very low) will be made by two review authors working independently, with disagreements resolved by discussion. Judgements will be justified, documented, and incorporated into reporting of results for each outcome.

We plan to extract study data, format our comparisons in data tables and prepare a summary of findings table before writing the results and conclusions of our review.

- The SoF for the main comparison will appear at the front of the published review
- Additional SoFs can be prepared for other important comparisons (those reported in full in the abstract) but it is not essential to have an SoF for every comparison
- Include a maximum of seven outcomes on each SoF
- Include the same outcomes for each comparison; include primary outcomes and adverse effects.
- The same comparisons and outcomes should be reported in the abstract as in the SoF

PART E – Results

Description of studies

Results of the search

The search retrieved 97 articles. Ten studies (12 articles) were potentially eligible and were retrieved in full text. Three studies (four articles) met our inclusion criteria. We excluded five studies and two are ongoing. See study tables: Characteristics of included studies, Characteristics of excluded studies, Characteristics of studies awaiting classification.

- Include a PRISMA flow chart – insert as a figure

Included studies

- No text should appear directly under this heading. Insert four subheadings (as below) and briefly summarise important points. Include full details of individual studies in Characteristics of included studies table (not this section).

Study design and setting

Three parallel-design randomised controlled trials (RCTs) were included. All were multicentre studies conducted in ART clinics in the Netherlands or the UK.

Participants

The studies included 811 subfertile women undergoing ART. Their mean age ranged across studies from 32 to 37 years.

Interventions

All three studies compared Treatment A with Treatment B

Outcomes

All three studies reported live birth and multiple pregnancy

All three studies also reported other adverse effects, but none included data suitable for analysis

Excluded studies

Five studies were excluded from the review, for the following reasons:

- *3/5 were not RCTs*
- *2/5 did not compare Treatment A versus Treatment B*

- Only studies that initially appeared eligible should be listed as excluded studies (i.e. if you had to read the full-text publication in order to determine that it was ineligible)

Risk of bias in included studies

Allocation (Selection bias)

Address both generation of random sequence and allocation concealment under this

heading. Use separate paragraphs or subheadings.

Sequence generation

One study was rated as at low risk of selection bias related to sequence generation, as it used computer randomisation or a random numbers table. The other two studies did not describe the method used and were rated as at unclear risk of this bias.

Allocation concealment

All studies failed to describe methods of allocation concealment and we rated these as at unclear risk of bias for this domain.

Blinding of participants and personnel (Performance bias)

Blinding of outcome assessors (Detection bias)

We did not consider that blinding was likely to influence findings for the primary review outcomes (live birth and multiple pregnancy). Blinding might influence outcomes for other adverse events, but no studies reported relevant data for this outcome.

- Consider the degree to which blinding is likely to influence specific outcomes.

Incomplete outcome data (Attrition bias)

All three studies analysed all or most (>95%) of the women randomised and we judged them to be at low risk of attrition bias.

Selective reporting (Reporting bias)

We rated all three studies as at low risk of selective reporting bias. All outcomes planned in the protocols were reported and these included live birth and multiple pregnancy (i.e. the most clinically relevant outcomes).

Other potential sources of bias (Other bias)

In one study there was a statistically significant baseline difference in age between the two groups and the risk of bias was deemed unclear. We found no potential sources of within-study bias in the other two studies.

- Do not include funding source, power calculations or ethics approval in this section, as they do not affect internal validity. These issues should be reported in the Characteristics of Included Studies table. If issues such as funding are of concern, they can be reported in the Results section (Description of studies), and the Discussion section (Quality of the evidence).

Effects of interventions

Notes – authors please read and implement!

- Separate primary and secondary outcomes.
- Use the same order of comparisons and outcomes and numbering system as in your Methods section and data tables.
- Include all pre-specified comparisons and outcomes: if there are no relevant data, say so
- Use this format for presenting results: (RR 0.89, 95% CI 0.75 to 1.05, three RCTs, n=811, I^2=31%, low quality evidence).
- Report and interpret the units used for continuous outcomes (e.g. a VAS scale of 0-10 where 0 is pain-free and 10 is unbearable pain): report this in the Abstract, main text and the 'comment's section of the SoF table.
- Do not confuse lack of evidence of an effect with evidence of a lack of effect: say "There was no evidence of a difference' not "There was no difference"
- Avoid saying "There was no significant difference".
- Consider clinical rather than statistical significance
- If there are multi-arm studies, avoid double-counting of controls.
- Do not describe the results of individual studies unless there is only one study in the comparison.
- If presenting multiple sensitivity analyses or different ways of subgrouping the same studies, present these in summary form (e.g. a single Table or Figure) and not in multiple forest plots.
- Report all pre-specified sensitivity and subgroup analyses at the end of each comparison. If there were too few studies to conduct the analyses, state this.
- Interpret subgroups very cautiously. In general, do not report them in the abstract.
- Report any post-hoc analyses at the end of each comparison, noting that they were not pre-specified and that they require extra caution in interpretation.
- Report the results of funnel plots E.g. "Funnel plots for the primary outcomes (live birth and ongoing pregnancy) did not suggest reporting bias".
- If there were too few studies to construct a funnel plot, state this in the results section.
- Acknowledge any substantial statistical heterogeneity detected and explore it (e.g. by means of sensitivity analyses).
- Explain the effect estimates for important clinical outcomes in a user-friendly way. We suggest using percentages (as in the example above), derived from the Summary of findings table.

How to format your findings:

1. Comparison of Treatment A versus Treatment B

Primary outcomes

1.1 Live birth

There was no conclusive evidence of a difference in cumulative live birth rate between Treatment A and Treatment B (RR 0.89, 95% CI 0.75 to 1.05, three RCTs, n=811, I2=0%, low quality evidence). This suggests that if the chance of live birth following Treatment A is assumed to be 42%, the chance following Treatment B would be between 31% and 44%.

There were too few studies to conduct any planned sensitivity analyses.

1.2 Multiple pregnancy

Treatment A was associated with lower multiple pregnancy rates than Treatment B (RR 0.04, 95% CI 0.01 to 0.15, three RCTs, n = 811, I2 = 23%, low quality evidence). This suggests that if the risk of multiple pregnancy following Treatment B is assumed to be 13%, the risk following Treatment A would be between 0% and 2%.

Secondary outcomes

1.3 Other adverse events

No studies reported on other adverse events

2*. Comparison of Treatment B versus Treatment C)*

Primary outcomes

2. 1 Live birth

2.2 Multiple pregnancy

Secondary outcomes

2.3 Other adverse events

PART F - Discussion

No text should appear directly under this heading

Summary of main results

- Briefly summarise the main review findings, directly addressing the objectives. Highlight and outstanding uncertainties, balancing important benefits against important harms. Express results in the most consumer-friendly way possible. Refer to quality of evidence (from SoF table)

Overall completeness and applicability of evidence

This section addresses the external validity of the review.

- Did the included studies answer the review question?
- Were relevant participants, interventions and outcomes investigated?
- Do the review findings support current practice?
- Comment on studies that measured outcomes but had no 'usable' data

Quality of the evidence

This section addresses the internal validity of the review.

- How robust are the conclusions?
- Use the GRADE ratings from the Summary of Findings (SoF) table to describe the quality of the evidence for each comparison, and use the footnotes from the SoF table to summarise the limitations of the evidence.
- Discuss limitations of the review at study and outcome level (e.g. regarding risk of bias), and at review-level (e.g. incomplete identification of studies, reporting bias).

Potential biases in the review process

Comment on the strengths and limitations of the review process

- *Were all relevant studies identified?*
- *Could review authors' methods have introduced bias?*

Agreements and disagreements with other studies or reviews

How do the review findings fit into the wider research context?

PART G Authors' conclusions

Implications for practice

- We suggest making this identical to the conclusions in the abstract
- Do not go beyond the evidence reviewed, mention GRADE ratings
- If relevant, summarise the likely benefits and risks/costs of the intervention and for whom it should be considered.

Implications for research

- Which questions have been well answered (no further trials needed)
- Which questions remain unanswered (further trials needed)
- Whether further trials in selected populations are warranted
- Identify any new research areas (dose modification, combined therapies etc).
- If recommending further research, structure the implications for research to address the nature of evidence required, including population, intervention comparison, outcome, and type of study.

Differences between protocol and review

- If planning any departures from Methods specified in protocol, check with editorial base first
- If changes are approved, document and justify the changes in the Differences from Protocol section.