PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Effects of oil-soluble versus water-soluble contrast media at hysterosalpingography on pregnancy outcomes in women with a low risk of tubal disease : study protocol for a randomized controlled trial
AUTHORS	Liang, Guiling; Zhu, Qian; He, Xiaoqing; Wang, Xiaofeng; Jiang, Ling; Zhu, Chenfeng; XIE, LI; Qian, Zhaoxia; Zhang, Jian

VERSION 1 – REVIEW

REVIEWER	Rui Wang Monash University, Australia
REVIEW RETURNED	28-Apr-2020

GENERAL COMMENTS	Liang and colleagues presented a trial protocol aiming to compare the effectiveness of oil-soluble contrast medium (OSCM) and water-based contrast medium (WSCM) during HSG in women at low risk of tubal pathology. According to the protocol, this is a conformation trial that repeats the H2Oil trial from the Netherlands (Deyer 2017) in a Chinese setting, but contrast media from different manufacturers will be used. The reporting of the protocol does not reach the required publishing standard in its current form. I have the following major and minor concerns. Major comments: - There are some inconsistencies between the protocol and trial registration (ChiCTR2000031612). The trial registration only included seven outcomes and all the “Measure time point of outcome” were left blank. Therefore, it is unclear from the trial registration that the trial has a 36-month follow-up period. - The follow-up of the trial will be 36 months and outcomes will be assessed multiple times at 6, 12, 24 and 36 months (Figure 1, table 1 and page 11). If this is correct, it is unclear how these outcomes at different time points will be analysed and whether multiplicity will be considered. In addition, please also clearly state when the outcomes will be measured in “secondary outcome measures” section. - The authors simply used the ongoing pregnancy rate in the WSCM group from Dreyer 2017 for the sample size calculation without further justification. It is unclear whether the baseline risk in the setting is similar to that in the Netherlands. The authors included a 10% additional sample size to account for lost-to follow up, but 934 *1.1= 1028 (not 1040 as the authors stated). Did the sample size calculation account for the adjustment for interim analysis? - Many key items in the SPIRIT checklist did not seem to be reported and quite some boxed in the SPIRIT checklists was not ticked. - Will be authors prepare a full statistical analysis plan? If not, some details on the analysis should be addressed in the protocol. Is there any statistical/methodologist in the team that will lead the analysis? Detailed comments: Title: “in the diagnosis of tubal patency” is misleading as this trial

	does not address diagnostic test accuracy at all – please remove to avoid confusion. Please try to incorporate the study population of interest in the title. Also please use the same terminology across the protocol, i.e. either oil-based contrast or oil-soluble contrast. Abstract: - Introduction: Please address the evidence gap in a direct and concise way. “The majority of infertility is caused by tubal” is confusing as this trial does not include any women with tubal infertility. - Methods and analysis: Please add the time point for the primary outcome, i.e. at 6 months after randomisation. No sample size calculation or statistical analyses were mentioned here. Strength and limitations: - The first bullet is the objective, not a strength. - The second bullet – as previous cohort studies have reported thyroid functions, the authors may wish to add “in an RCT”. - The long-term follow-up could be listed as a strength Introduction Page 5, line 20. No reference here. The prevalence of infertility in China was higher in other national reports from China (Zhou 2017). Line 23. Reference 2 is incorrect as it does not mention any data on tubal pathology from China. Tubal pathology is irrelevant here as this trial does not include women with tubal pathology. Line 34. “In China, it is the preferred method to check tubal patency.” Please provide a reference here. HyCoSy and HyFoSy are also commonly used in many fertility clinics in China. Line 44. Why specifically mentioning “Ultra-fluid lipiodol” here? This product is not the intervention used in this trial. Instead, the general features of all OSCM should be introduced here. Page 5 Line 17 “numerous studies” Please provide references. Page 6 Line 4- line 35. This paragraph should be more focused. Evidence from relevant systematic reviews should be used to identify the evidence gap. Page 6 Line 50. This paragraph is a repeat of the objective /hypothesis. Please remove. Page 7 Line 10. If the investigation of thyroid dysfunction is important, the authors could consider adding it as a secondary objective of this trial. Methods and analysis: Page 7 Line 57. It is crucial to define low risk for tubal pathology as this is the most important feature of the inclusion criteria. Page 8 Line 25 “a post-wash total motile sperm count < 3 x10^6 spermatozoa/ml” This is also inconsistent with the protocol. Will the screening include a post-wash sperm count for every participant? Page 8 Line 45. “Prior to randomization, clinical data will be entered in the digital platform.” It is unclear what the authors mean here as data, except for those on screening details, are usually entered after randomisation. Page 8 Line 57. Please describe the intervention (OSCM) before the control group (WSCM). Page 8 Line 23. The last sentence “Visual analogue and qualitative scales ...” belongs to the outcomes. Page 9 Line 27. The “Withdrawal of participants” section is too repetitive. Please consider condensing this section into a short paragraph. Page 10 Table 1. What is the difference between “safety inf intervention”, “side effects/complication” and “adverse event review and evaluation”? Does “assisted reproductive technology” include IUI as well? It is unclear from the table whether baseline pain and thyroid function were measured before the intervention.

	Page 10 Line 53. Do you mean Day 1 of the menstrual cycle? Why not just before HSG? Page 11 Line 13. Operation – do you mean laparoscopic /hysteroscopic surgery? Page 11. Secondary outcomes. Please clearly define all the secondary outcomes, including when they will be measured. Page 11 Line 21. Will the same biomarkers for thyroid function tested for the participants and their newborns? Page 11 Line 42. Pregnancy leading to live birth – do you consider it as a dichotomous outcome or time-to-event outcome? If the former, it is the same as pregnancy; if the latter, please rephrase as time to pregnancy leading to live birth. Line 51. Please list the biomarkers for thyroid function. Line 54. Please remove “between two groups” Page 12 Line 33. Please consider removing Figure 2 as it is not informative. Line 39. The first paragraph is repetitive – consider removing. Line 44 onwards. Please use the future tense. Line 57. Will baseline information be adjusted for the analysis of continuous outcomes? Line 60 “time to pregnancy resulting in an ongoing pregnancy” does not seem to be an outcome of this study. Page 14 Line 16. Who (Clinicians/statisticians) will be the potential DMC member? Discussion: Please revise in line with the comments on “Strength and limitations” above. Ethics and dissemination: This section was not reported. Please also add relevant information on trial status. References Dreyer et al. Oil-Based or Water-Based Contrast for Hysterosalpingography in Infertile Women. N Engl J Med. 2017 25;376(21):2043-2052. Zhou et al. Epidemiology of infertility in China: a population-based study. BJOG. 2018;125(4):432-441.

REVIEWER	koks CAM Maxima Medical centrum The Netherlands
REVIEW RETURNED	03-May-2020

GENERAL COMMENTS	1. change the title, because the focus in the title is diagnostic performance of HSG instead of therapeutic effect. 2. prim outcome is ongoing pregnancy, sec outcome live birth rate, clinical pregnacy rate, miscarriage, ectopic pregnancy , pregnancy leading to live birth, pain scores, thyroid function of patient and neonate. Why are you using also clinical pregnacy-rate? Describe the sec outcomes in the same order in the protocol. 3. Use oil-versus water based contrast in the same order in the protocol instead of using it sometimes in a different order 4. page 6, line 33, investigate instead ofcheck 5. page 6, line 44, which adverse events?, also later in the protocol 6. SAE, and how are they handled 7. how are the patients given oral and written information, who is giving these information, gynaecologist, radiologist, researchnurse. Who performs the randomisation. Who is handling the information ? 8. page 8 line 54; inclusion if Chlam PCR is neg or no history of chlamydia infection . I think this should be no or but and no history of chlamydia-infection. How is this checked, by CAT (chlamydia-antibody-titer).

	9. Will the HSG be done under fluoroscopy ? When will the procedure stop? When there is intravasation? More than 10 ml of contrastmedia, other reasons. When will the painscores be done and how? 10. ITT, use first the full description. 11. what is the timeframe of the study, when will it start and when is it suposed to stop, how many HSG are done in the hospital . 12. Very good idea to measure thyroid function, but can the authors explain why they will monitor so often and so long? what do they expect? Why do they measure TSI and TPO-antibodies, and do they measure them also at 8 different time-points. Can they explain what kind of thyroid functions will be measured in the neonate I don't have the expertise to statistical review this protocol.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Thank you for spending time on reviewing our manuscript and for your valuable suggestions, which we believe have greatly helped us improve our manuscript.

Major comments:

- There are some inconsistencies between the protocol and trial registration (ChiCTR2000031612). The trial registration only included seven outcomes and all the "Measure time point of outcome" were left blank. Therefore, it is unclear from the trial registration that the trial has a 36-month follow-up period.

Response: Thank you for your careful review. Our protocol is more detailed and accurate. We have registered the trial and published the protocol to ensure transparency of our clinical research.

- The follow-up of the trial will be 36 months and outcomes will be assessed multiple times at 6, 12, 24 and 36 months (Figure 1, table 1 and page 11). If this is correct, it is unclear how these outcomes at different time points will be analysed and whether multiplicity will be considered. In addition, please also clearly state when the outcomes will be measured in "secondary outcome measures" section.

Response: We are sorry for the ambiguity of the secondary outcome measures. We have revised the part of "***secondary outcome measures".*** In this section, we state the time points of the secondary outcome measures. (Please see Page 12, Line 15-22 and Page 13, Line 1-18)

In our study, the follow-up of secondary outcomes is considered as exploratory outcomes to support the results concerning the primary outcome and will help to generate new hypotheses for future research. We will not draw any firm conclusions from the secondary outcomes. Hence, we may not focus on the multiplicity of type I error on secondary outcomes.\

The revision is as follows:

Secondary outcome measures

1. The thyroid function of patients will be examined before HSG, and at 4, 8, 12, and 24 weeks, and 9–12 months after HSG. We will determine free triiodothyronine (FT3), free thyroxine (FT4), thyroid stimulating hormone (TSH), antithyroglobulin antibodies (TgAb), and antithyroid peroxidase antibodies (TPOAb) levels at different time-points [26,27].
2. The thyroid function of the neonates will be tested within 3 to 7 days after birth. TT4 (total thyroxine), FT4, and TSH will be detected [27].
3. Pain scores during hysterosalpingography will be measured by means of the Visual-Analogue Scale for Pain (scores range from 0.0 to 10.0 cm, with higher scores indicating more severe pain). The pain scores will be recorded by a trained nurse. Meanwhile, the time from the beginning of the contrast injection to the occurrence of pain will be recorded.
4. The rates of live birth rate/ clinical pregnancy / miscarriages/ ectopic pregnancy/ pregnancy leading to live birth will be assessed. Live birth is defined as the birth of at least one living child; clinical pregnancy is defined as an ultrasound visible gestational sac; miscarriage is defined as a spontaneous loss of pregnancy; ectopic pregnancy is defined as an embryo implanted outside the uterine cavity. Pregnancy leading to live birth is defined as the ratio of live births to clinical pregnancies. Each occurrence of one of these events will be recorded during the three-year follow-up.
5. The time to pregnancy resulting in an ongoing pregnancy is defined as the time from randomization to the first day of the last menstrual period plus 4 weeks. It will be considered when ongoing pregnancy occurs.
6. The costs and cost-effectiveness of OSCM / WSCM and assisted reproductive technology (ART) treatments in Chinese context.
7. The side effects or complications after therapy in both groups.

- The authors simply used the ongoing pregnancy rate in the WSCM group from Dreyer 2017 for the sample size calculation without further justification. It is unclear whether the baseline risk in the setting is similar to that in the Netherlands. The authors included a 10% additional sample size to account for lost-to follow up, but 934 *1.1= 1028 (not 1040 as the authors stated). Did the sample size calculation account for the adjustment for interim analysis?

Response: Thank you for this concern.We have reviewed some papers of studies on HSG in Chinese women. We found that the baseline information (such as age, BMI, the percentage of primary infertility) in Chinese infertile women is similar to that in the population in Dreyer's research. Therefore, we used the ongoing pregnancy rate in the WSCM group from Dreyer 2017 for the sample size calculation without further justification. (Reference: Yiqing T, ShilinZh, Wenfeng L, et al. Ethiodized poppyseed oil versus ioversolfor image quality and adverse events inhysterosalpingography: a prospectivecohort study. BMC Medical Imaging2019;19(1):50-7; Zijun D. Study on Clinical Effect of Ioversol and Iodized Oil in the HysterosalpingographyChina &Foreign Medical Treatment. 2018;36:175-7; Yanyan H, Jinglan L. Influence factors of pregnancy rate after selective salpingography and fallopian tube recanalization.Chinese Journal of Clinical Medicine 2019;26(1):62-4.)

The sample size calculation is 934/0.9=1037.7, so we will enrol 1040 patients. The sample size calculation did not account for the adjustment for interim analysis

- Many key items in the SPIRIT checklist did not seem to be reported and quite some boxed in the SPIRIT checklists was not ticked.

Response: We are sorry for the incompleteness of the SPIRIT checklist. We have completed the checklist.(Please see the file SPIRIT checklist).We have also supplemented the key items in the protocol.

- Will be authors prepare a full statistical analysis plan? If not, some details on the analysis should be addressed in the protocol. Is there any statistical/methodologist in the team that will lead the analysis?

Response: Thank you for the suggestion. We have addressed the frame of statistical analysis in the protocol.(Please see Page 14, Line 21-22 and Page 15, Line 1-15)The detailed analysis will be done by an independent statistician before interim analysis.

Detailed comments:

Title: "in the diagnosis of tubal patency" is misleading as this trial does not address diagnostic test accuracy at all – please remove to avoid confusion. Please try to incorporate the study population of interest in the title. Also please use the same terminology across the protocol, i.e. either oil-based contrast or oil-soluble contrast.

Response: We apologize for the confusion. We have amended the title as "Effects of oil-soluble contrast versus water-soluble contrast media at hysterosalpingography on pregnancy outcomes in Chinese women: study protocol for a randomized controlled trial".(Please see Page 1, Line 1-4)We have changed allinconsistent terminology across the protocol.

Abstract:

- Introduction: Please address the evidence gap in a direct and concise way. "The majority of infertility is caused by tubal" is confusing as this trial does not include any women with tubal infertility.

Response: Thank you for your suggestion, and we are sorry for the confusion. We have removed the sentence about tubal infertility. We have made other revisions in the paragraph.(Please see Page 2, Line 2-9)

The revision is as follows:

In recent years, due to various factors, the infertility rate in China has increased and now affects more than 10% of women of reproductive age. Hysterosalpingography (HSG) is a common diagnostic procedure during fertility examinations. However, there is no consensus on the choice of contrast agents and their effects. As the largest multi-centre randomised controlled trial (H2Oil trial from the Netherlands) has shown the oil contrast at HSG can enhance the fertility, we propose this study to

answer the question of whether the use of oil-soluble contrast media results in increased pregnancy rates in Chinese women undergo HSG.

- Methods and analysis: Please add the time point for the primary outcome, i.e. at 6 months after randomisation. No sample size calculation or statistical analyses were mentioned here.

Response: We apologize for the incompleteness. We have added the time point for the primary outcome and sample size calculation.(Please see Page 2, Line 13-18)

The revision is as follows:

To evaluate the potential superiority of OSCM group, with 1:1 allocation ratio, 90% statistical power and a two-sided significance level of 5%, and considering 10% lost-to-follow up or protocol violation, we have calculated a sample of 520 women per group for a total of 1040 would need to be enrolled. The primary outcome is the rate of ongoing pregnancy during 6 months after randomisation.

Strength and limitations:

- The first bullet is the objective, not a strength.

- The second bullet – as previous cohort studies have reported thyroid functions, the authors may wish to add "in an RCT".

- The long-term follow-up could be listed as a strength

Response: Thank you for the suggestion. We have revised the strengths and limitations section.(Please see Page 3, Line 8-18)

Introduction

Page 5, line 20. No reference here. The prevalence of infertility in China was higher in other national reports from China (Zhou 2017).

Response: Thank you for your careful reviewing; we have added the reference.(Please see Page 4, Line 5 and the reference 2)

The reference is “Zhenwu Z, WenliL. Estimating the Prevalence of Infertility in ChinaUsing Census Data. Population Research 2020;44(02):3-17.”

Line 23. Reference 2 is incorrect as it does not mention any data on tubal pathology from China. Tubal pathology is irrelevant here as this trial does not include women with tubal pathology.

Response: We appreciate your meticulous reviewing. We have studied the reference again and found that we cited an incorrect reference; therefore, we removed this sentence because of its irrelevance.(Please see Page 4, Line 5)

Line 34. “In China, it is the preferred method to check tubal patency.” Please provide a reference here. HyCoSy and HyFoSy are also commonly used in many fertility clinics in China.

Response: We are sorry for our neglecting on the reference. We have listed the reference in the protocol.(Please see Page 4, Line 10-11 and reference 6)

The reference is “Jinxia B, Jing H, Shengli W. Clinical Progress of Tubal Patency Tests.J Int ObstetGynecol2020;47:111-4. ”

Line 44. Why specifically mentioning “Ultra-fluid lipiodol” here? This product is not the intervention used in this trial. Instead, the general features of all OSCM should be introduced here.

Response: We are sorry for the improper description. We have revised the expression and add the general features of all OSCM.(Please see Page 4, Line 15-17)

The revision is as follow:

Oil-soluble contrast media (OSCM) is represented by iodinated oil injection, which uses poppy seed oil as the raw material. It provides clear images and may have some anti-inflammatory effects that perhaps enhance fertility.

Reference: Zijun D. Study on Clinical Effect of Ioversol and Iodized Oil in the Hysterosalpingography. China &Foreign Medical Treatment 2018;12:175-7.

Page 5 Line 17 "numerous studies" Please provide references.

Response: Thank you for pointing this out. We have listed the reference in the protocol.(Please see Page 5, Line 15-16 and reference 18-21)

The references are as follows:

Alper MM, Garner PR, Spence JE, et al. Pregnancy rates afterhysterosalpingography with oil- and water-soluble contrast media. ObstetGynecol1986;68:6–9.

de Boer AD, Vemer HM, Willemsen WN, et al. Oil or aqueous contrastmedia for hysterosalpingography: a prospective, randomized, clinical study.Eur J ObstetGynecolReprod Biol 1988;28:65–8.

Lindequist S, Rasmussen F, Larsen C. Use of iotrolan versus ethiodized poppy-seed oil in hysterosalpingography. Radiology,1994,191:513–517.

Spring DB, Barkan HE, Pruyn SC. Potential therapeutic effects of contrast materials in hysterosalpingography: a prospective randomized clinical trial.Kaiser Permanente Infertility Work Group. Radiology,2000;214:53–57.

Page 6 Line 4- line 35. This paragraph should be more focused. Evidence from relevant systematic reviews should be used to identify the evidence gap.

Response: Thank you for your suggestion. We have revised this paragraph.(Please see Page 5, Line 6-9)

The revision is as follows:

Fang's systematic review showed that HSG using OSCM may promote the ongoing pregnancy rate through a comprehensive analysis of six studies. However, the review did not include Chinese population and did not pay close attention to the thyroid function or long-term effects

Reference: Fang F, Yu B, Yu Z, et al. Oil-based versus water-based contrast for hysterosalpingography in infertile women: a systematic review and meta-analysis of randomized controlled trials. Fertility & Sterility 2018:S0015028218302760-.

Page 6 Line 50. This paragraph is a repeat of the objective /hypothesis. Please remove.

Response: We thank you for the suggestion. We have revised the paragraph.(Please see Page 6, Line 7-9)

The revision is as follows:

In summary, there is no consensus on the choice of contrast agents and their effects. In view of this uncertainty, we plan a single-centre, randomized, controlled, parallel group, superiority trial in infertile Chinese women with a low a priori chance of tubal pathology.

Page 7 Line 10. If the investigation of thyroid dysfunction is important, the authors could consider adding it as a secondary objective of this trial.

Response: We agree with your opinion and have listed the thyroid function as a secondary objective of this trial.(Please see Page 12, Line 16-22)

Methods and analysis:

Page 7 Line 57. It is crucial to define low risk for tubal pathology as this is the most important feature of the inclusion criteria.

Response: Thank you for the suggestion. We have supplemented the definition of low risk for tubal pathology.(Please see Page 7, Line 18-22)

The definition is as follows:

The patient has not been exposed to high risk factors of tubal pathology, such as history of chlamydia infection, pelvic inflammatory disease, known endometriosis or adenomyosis, pelvic abdominal surgery [including salpingostomy or salpingectomy for ectopic pregnancy and complicated appendectomy] and/or peritonitis.

Reference:Coppus SF, Verhoeve HR, Opmeer BC, et al. Identifying subfertileovulatory women for timely tubal patency testing: a clinical decision rule based onmedical history. Hum Reprod 2007;22(10):2685-92.

Page 8 Line 25 "a post-wash total motile sperm count < 3 x10^6 spermatozoa/ml" This is also inconsistent with the protocol. Will the screening include a post-wash sperm count for every participant?

Response: Thank you for this concern. We have added relevant context in the inclusion criteria.(Please see Page 7, Line 14-15)In our hospital, a couple comes to see a doctor for infertility together; therefore, the screening will include a post-wash sperm count for every participant.

Page 8 Line 45. "Prior to randomization, clinical data will be entered in the digital platform." It is unclear what the authors mean here as data, except for those on screening details, are usually entered after randomisation.

Response: Thank you for your careful review. We mean that screening data will be entered in the digital platform prior to randomization. We have modified the sentence in the protocol.(Please see Page 9, Line 1-2)

Page 8 Line 57. Please describe the intervention (OSCM) before the control group (WSCM).

Response: Thank you for your careful review. We have changed the order of the description.(Please see Page 9, Line 14-19)

Page 8 Line 23. The last sentence "Visual analogue and qualitative scales ..." belongs to the outcomes.

Response: Thank you for this suggestion. We have removed the sentence.

Page 9 Line 27. The "Withdrawal of participants" section is too repetitive. Please consider condensing this section into a short paragraph.

Response: Thank you for the suggestion. We have simplified this paragraph.(Please see Page 10, Line 15-21)

The revision is as follows:

Participants will be free to withdraw from the study at any time upon request. An investigator may discontinue or withdraw a participant from the study for the following reasons: the participant meets an exclusion criterion; violation of research protocol; the patient is experiencing an urgent medical situation.

The reasons for participant discontinuation or withdrawal from the study will be recorded on the Case Report Form (CRF). Data collection will continue if data can be safely acquired, and the data will be used for the ITT (intention-to-treat) analyses.

Page 10 Table 1. What is the difference between "safety inf intervention", "side effects/complication" and "adverse event review and evaluation"? Does "assisted reproductive technology" include IUI as well? It is unclear from the table whether baseline pain and thyroid function were measured before the intervention.

Response: We apologize for the repetition and ambiguity of the context. We have deleted the "side effects/complication" and "adverse event review and evaluation" in the table. ART includes IUI as well. The thyroid function should be measured 7 days before the intervention to exclude thyroid dysfunction and record the basic thyroid function.(Please see Page 11-12 and Table 1)As the pain in our study is caused by the operation of HSG, we think the baseline pain of the patients is scored 0 (VAS scores).

Page 10 Line 53. Do you mean Day 1 of the menstrual cycle? Why not just before HSG?

Response: We are sorry for the unclear expression. Day 1 means the date of HSG.We have clarified this.(Please see Page 11, Line 6)

Page 11 Line 13. Operation – do you mean laparoscopic /hysteroscopic surgery?

Response: We are sorry for the unclear expression. Operations include laparoscopic and hysteroscopic surgery. We have added the description in the protocol.(Please see Page 12, Line 5-6)

Page 11. Secondary outcomes. Please clearly define all the secondary outcomes, including when they will be measured.

Response: We thank you for the suggestions. We have revised the section of “Secondary outcome measures”. (Please see Page 12, Line 15-22 and Page 13, Line 1-18)

Page 11 Line 21. Will the same biomarkers for thyroid function tested for the participants and their newborns?

Response: We are sorry for the neglecting on the details. The biomarkers for the thyroid function for the participants and their newborns are not exactly the same. In patients,FT3, FT4, TSH, TgAb, and TPOAb level will be tested in different time points. However, in newborns, TT4, FT4, and TSH will be assessed 3 to 7 days after birth.(Please see Page 12, Line 16-22)

Reference:Ad Hoc Writing Committee for Guideline on Diagnosis and Management of Thyroid Diseases during Pregnancy and Postpartum (2nd edition);Chinese Society of Endocrinology, Chinese Medical Association; Chinese Society of Perinatology, Chinese Medical Association. Guideline on Diagnosis and Management of Thyroid Diseases during Pregnancy and Postpartum (2nd edition). 2019;22(8):505-39.

Page 11 Line 42. Pregnancy leading to live birth – do you consider it as a dichotomous outcome or time-to-event outcome? If the former, it is the same as pregnancy; if the latter, please rephrase as time to pregnancy leading to live birth.

Response: We are sorry for the unclear description. We consider the pregnancy leading to live birth as a dichotomous outcome. Pregnancy leading to live birth is defined as the ratio of live birth to clinical pregnancy. We attempt to be consistent with the H2Oil trial to make the data comparable.(Please see Page 13, Line 10-11)

Line 51. Please list the biomarkers for thyroid function.

Response: Thank you for this suggestion. We have listed the biomarkers for thyroid function in the protocol.(Please see Page 12, Line 16-22)

Line 54. Please remove "between two groups"

Page 12 Line 33. Please consider removing Figure 2 as it is not informative.

Line 39. The first paragraph is repetitive – consider removing.

Line 44 onwards. Please use the future tense.

Response: Thank you for your careful reviewing. We have made the revisions as you have suggested.(Please see Page 12, Line 16-20; Page 14, Line 19; Page 14, Line 13and Page 15, Line 3)

Line 57. Will baseline information be adjusted for the analysis of continuous outcomes?

Response: Wewill not adjust the baseline information for the analysis of continuous outcomes***.***

Line 60 "time to pregnancy resulting in an ongoing pregnancy" does not seem to be an outcome of this study.

Response: Thank you for your careful review. We have added the time to pregnancy resulting in an ongoing pregnancy as a secondary outcome.(Please see Page 13, Line 15-17)

Page 14 Line 16. Who (Clinicians/statisticians) will be the potential DMC member?

Response: Thank you for your question. The ethics committee that approved the protocol and an Independent Data Monitoring Committee (IDMC) will be the members of DMC.(Please see Page 15, Line 18-20)

Discussion: Please revise in line with the comments on "Strength and limitations" above.

Response: Thank you for your suggestion and we have revised the discussion.(Please see Page 17, Line 13-18)

The revision is as follows:

The strengths of this study are as follows: First, this will be the first study to assess thyroid function of patients at different time points who undergo HSG and assess their newborns' thyroid function in a RCT. Second, compared to other studies, we will extend the follow-up period to evaluate the long-term impacts of HSG using different reagents. Finally, this will be the first record of pain intensity during HSG and the time interval of when pain occurs.

Ethics and dissemination: This section was not reported.

Response: We apologize for the omission of this section. We have added it in the protocol. (Please see Page 16, Line 18-22 and Page 17, Line 1-2)

Please also add relevant information on trial status.

Response: Thank you for your review. We have added relevant information on trial status in the protocol. (Please see Page 18, Line 4-6)

The recruitment has not started yet. We are going to begin the recruitment in June 2020 and expect to complete the recruitment in June 2021.

References

Dreyer et al. Oil-Based or Water-Based Contrast for Hysterosalpingography in Infertile Women. N Engl J Med. 2017 25;376(21):2043-2052.

Zhou et al. Epidemiology of infertility in China: a population-based study. BJOG. 2018;125(4):432-441.

Reviewer 2

We thank the reviewer for these constructive comments, which we believe made our manuscript more logical and well-organized.

1. ***change the title, because the focus in the title is diagnostic performance of HSG instead of therapeutic effect.***

Response: Thank you for your suggestion. We have amended the title as “Effects of oil-soluble contrast versus water-soluble contrast media at hysterosalpingography on pregnancy outcomes in Chinese women: study protocol for a randomized controlled trial”.(Please see Page 1, Line 1-4)

2. prim outcome is ongoing pregnancy, sec outcome live birth rate, clinical pregnancy rate, miscarriage, ectopic pregnancy , pregnancy leading to live birth, pain scores, thyroid function of patient and neonate. Why are you using also clinical pregnancy-rate? Describe the sec outcomes in the same order in the protocol.

Response: Thank you for your careful reviewing. We have changed the order of the secondary outcomes. Clinical pregnancy can reflect fertility of the patients to a certain degree. Moreover, we have designed this study to be consistent with the H2Oil trial to make the data comparable.

3. Use oil-versus water based contrast in the same order in the protocol instead of using it sometimes in a different order

Response: Thank you for the suggestion. We have made revisions to be parallel throughout.

4. page 6, line 33, investigate insteadofcheck

Response: Thank you for your careful reviewing. We have changed the verb at your suggestion. (Please see Page 4, Line10)

5. page 6, line 44, which adverse events?, also later in the protocol

Response: Thank you for the suggestion. We have added the adverse events in the protocol under the section “Safety assessment”. (Please see Page 13, Line20-22 and Page 14, Line 1-9)

The context is as follows:

All adverse events (AEs) will be recorded during theentire study period. An adverse event is defined as an event during or following hysterosalpingography or follow-up which was not intended to happen and is suspected to be a complication of the intervention performed. Common adverse events of HSG

include allergic reactions, artificial abortion syndrome, abdominal pain, and intravasation. The severity of the AEs will be primarily rankedas ‘mild’, ‘moderate’, or ‘severe’. Thecauses of the AEs will be rated as ‘definitely related’, ‘probably related’, ‘possibly related’, ‘probably not related’, ‘definitely not related’, or ‘unknown’.

A severe adverse event (SAE) is defined as death, illness necessitating hospitalization, disability, or congenital malformation. All SAEs will be reported to the ethics committee that approved the protocol within 24 hours.

6. SAE, and how are they handled

Response: Thank you for this concern. We have added a section on SAE to the protocol.(Please see Page 14, Line 7-9)All SAEs will be reported to the ethics committee that approved the protocol within 24 hours.

7. how are the patients given oral and written information, who is giving these information, gynaecologist, radiologist, researchnurse.

Who performs the randomisation. Who is handling the information ?

Response: Thank you for your question. After tubal testing is indicated and planned, the informed consent for screening patients will be performed by a gynaecologist in outpatient services, including possible risks and benefits. The informed consent process is intended to help patients make a decision on whether or not they want to participant in this clinical trial. Oral and written informed consent will be obtained.(Please see Page 7, Line 2-6) Randomisation will be performed by an independent statistician using a web-based Research Electronic Data Capture (REDCap) system. (Please see Page 8, Line 20-22)The investigator will record the information in the CRFs and source documents.(Please see Page 11, Line 3)

8. page 8 line 54; inclusion if Chlam PCR is neg or no history of chlamydia infection . I think this should be no or but and no history of chlamydia-infection. How is this checked, by CAT (chlamydia-antibody-titer).

Response: Thanks for your careful reviewing. We have used ‘and’ instead of ‘or’. Chlamydia-infection was detected by vaginal secretion culture through Chlamydia antigen detection in our hospital.(Please see Page 7, Line 16-17)

9. Will the HSG be done under fluoroscopy ? When will the procedure stop? When there is intravasation? More than 10 ml of contrastmedia, other reasons.

Response: Thank you for this question. TheHSG will be performed under fluoroscopy. The procedure will stop when the uterine cavity was filled with contrast and the Fallopian tube development appears. Early intravasation into uterine and ovarian veins or lymphatics manifests as multiple thin ascending beaded channels on radiography. The amount varies with uterine size and peritoneal spillage; 10 to 30 mL is a typical dose. When patients undergo HSG in our hospital, up to 10 mL of contrast medium will be used.

Reference: Susanna I Lee, Aoife Kilcoyne. Hysterosalpingography. UpToDate 2019.

When will the painscores be done and how?

Response: Thank you for your question. Pain scores during hysterosalpingography are measured by means of the Visual-Analogue Scale for Pain (scores range from 0.0 to 10.0 cm, with higher scores indicating more severe pain). The pain scores will be recorded by a trained nurse. Meanwhile, the time from the beginning of the contrast injection to the occurrence of pain will be recorded.(Please see Page 13, Line 1-5)

10. ITT, use first the full description.

Response: Thank you for the suggestion. We have revised the description.(Please see Page 10, Line 21)

11. what is the timeframe of the study, when will it start and when is it suposed to stop, how many HSG are done in the hospital .

Response: Thank you for this question. We are going to begin the recruitment in June 2020, and we plan to take 1 year to recruit and 3 years for follow-up. In our hospital, there are 2500 to 4000 HSGs each year.

12. Very good idea to measure thyroid function, but can the authors explain why they will monitor so often and so long? what do they expect? Why do they measure TSI and TPO-antibodies, and do they measure them also at 8 different time-points.

Can they explain what kind of thyroid functions will be measured in the neonate

Response: Thank you for your acceptance of our idea. In Terumi's (2015) research, the mean level of TSH significantly increased at 4, 8, 12, and 24 weeks post-HSG compared with pre-HSG, and the mean value of FT3 and FT4 showed no significant difference at any of the time points compared with pre-HSG. However, it is a prospective observational study with only 22 enrolled patients. TSI and TPO-antibodies may be related to thyroiditis. Therefore, we would like to measure all biomarkers related to thyroid function at different time-points to explore the possible effects of contrast on thyroid.

The thyroid function of neonates will be tested 3 to 7 days after birth. TT4, FT4, and TSH will be detected.

Reference:

Kaneshige T, Arata N, Harada S, et al. Changes in Serum Iodine Concentration, Urinary Iodine Excretion and Thyroid Function After Hysterosalpingography Using an Oil-Soluble Iodinated Contrast Medium (Lipiodol). J Clin Endocrinol Metab 2015;100(3):E469–72.

Ad Hoc Writing Committee for Guideline on Diagnosis and Management of Thyroid Diseases during Pregnancy and Postpartum (2nd edition);Chinese Society of Endocrinology, Chinese Medical Association; Chinese Society of Perinatology, Chinese Medical Association. Guideline on Diagnosis and Management of Thyroid Diseases during Pregnancy and Postpartum (2nd edition). 2019;22(8):505-39.

I don't have the expertise to statistical review this protocol.

VERSION 2 – REVIEW

REVIEWER	Rui Wang Monash Univerity
REVIEW RETURNED	07-Jul-2020

GENERAL COMMENTS	Thanks for providing this R1. The protocol reads better. However, there are some remaining issues that need to be clarified. Major comments: 1. There are still several issues with the outcome definitions: - It is strange to set a time horizon of 6 months for the primary outcome (not live birth), but to have a time horizon of 36 months for all the other pregnancy outcomes. By the time of 36-month follow up, live birth resulting from pregnancies within the first 6 months should have been available. With the current description, I would assume that the authors plan to publish the primary report after all 3-year follow-up data are available. However, if the authors plan to publish the primary report when all data on pregnancy resulting from the first 6 months are available (including secondary pregnancy outcomes) and an additional report when 3-year follow-up data are available, they should make it clear. - There are still multiple inconsistencies on the descriptions of the secondary outcomes (methods, table 1 and figure 1). Figure 1 is really confusing as it indicates almost all outcomes will be evaluated 4 times (6, 12, 24, 36 months after randomisation), including thyroid function, which is inconsistent with table 1 and the texts in methods section. Given that all the outcomes have been presented in Figure 1, I wonder whether Figure 1 adds any useful information here. - “Pregnancy leading to live birth is defined as the ratio of live births to clinical pregnancies.” The intention of using pregnancy leading to live birth is for the analysis of time-to-event outcomes, and not for the analysis of dichotomous outcomes. It is poor methodological practice to use the clinical pregnancy as the denominator for the analysis of live birth. I would suggest considering this outcome as a time-to-event outcome (time to pregnancy leading to live birth) and replace “time to pregnancy resulting in ongoing pregnancy” by “time to pregnancy leading to live birth”. This also implies in the statistical analysis section. Please also clarify the time points at which all pregnancy outcomes are measured – only at 36 months or also at 6 months? 2. Sample size calculation. - The references used in the response letter to justify the sample size calculation included mostly women with PID, which is different from the trial population. The authors may wish to cite some data from their clinic or other papers on women with low risk of tubal disease to justify the sample size calculation. - The wording on lost to follow up is inconsistent with the way the authors calculated the data. “Additional 10%” would imply the total sample size =1.1* calculated sample size. The authors may wish to rephrase the last two sentences for sample size calculation (Page 14, Line 20-22) as “Thus, 1040 women (520 in each group) will need to be randomised, including 10% loss to follow-up or protocol variation.” 3. Statistical analysis. - The current statistical analysis is almost identical to the relevant paragraphs in Dreyer 2017, apart from the tense used. I would leave it to the editors to decide whether it is acceptable or not. - As the authors do not have a separate statistical analysis plan, they need to make this section more comprehensive according to the SPIRIT checklist, including any planned subgroup analysis,

	sensitivity analysis and the handling of missing data. Apart from intention to treat analysis, whether per protocol analysis is also planned. Some further detailed comments: Page 1 Title: When asked to "incorporate the population into the title", the authors should consider the most important characteristic of the population. In this case, the ethnicity does not seem to make a difference here. I would suggest replacing "Chines women" by "women with a low risk of tubal disease" (or pathology or something similar) if word limit allows. Please also remove the first "contrast" in the title. Page 2 Abstract: Introduction: Please incorporate the control group the objective. Methods: Please emphasize that the study population is women with a low risk of tubal pathology. Outcomes: Please refer to comments above on outcome definitions. Page 3 Strengths: Please do not overemphasize the "first" as this is neither the first study with long-term follow-up, nor the first study measuring pain. Long-term follow-up has been also performed in H2Oil trial. There have been many other trials reporting pain as outcome since 1990s, including the recent H2Oil Trial. Reporting long-term follow up itself is sufficient as a strength. Page 5, line 10: "did not pay close attention to" – did not report Page 6, line 4-5: As lipiodol is not the product of interest, "lipiodol" – OSCM Page7, line 17: As post-wash total motile sperm count is only available for couples undergoing IUI, I don't know what "post-wash total motile sperm count" refer to here. Page 10, line 5: "Consecutive"- Eligible Page 15, line 22. "Monitoring staff will consist of an Independent Data Monitoring Committee (IDMC) and an ethics committee". The authors still did not describe the composition of IDMC. How many members? Who are they (if not established yet, at least mentioning the number of clinical experts and number of statisticians)? Page 16, line 7. "An Independent Data Monitoring Committee (IDMC) and ethics committee will review data annually during the accrual period ….. At each meeting, the IDMC will be asked to give advice on whether the accumulated data from the trial, ….." As the trial is only planned to recruit patients for 12 months, it seems inappropriate to use "annually" "at each meeting" here. Page 18, line 8. Please update the trial status if applicable.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

We thank you for these constructive comments, which we believe made our manuscript more logical and well-organized.

Major comments:

1. There are still several issues with the outcome definitions:

- It is strange to set a time horizon of 6 months for the primary outcome (not live birth), but to have a time horizon of 36 months for all the other pregnancy outcomes. By the time of 36-month follow up, live birth resulting from pregnancies within the first 6 months should have been available. With the current description, I would assume that the authors plan to publish the primary report after all 3-year follow-up data are available. However, if the authors plan to publish the primary report when all data on pregnancy resulting from the first 6 months are available (including secondary pregnancy outcomes) and an additional report when 3-year follow-up data are available, they should make it clear.

Response: We are sorry for the ambiguity ofthe presentation of results. We plan to publish the primary report when all data on pregnancy resulting from the first 6 months are available (including secondary pregnancy outcomes) and an additional report when 3-year follow-up data are available. We have made relevant statements in the protocol.(Please see Page 24, Line 15-20)

- There are still multiple inconsistencies on the descriptions of the secondary outcomes (methods, table 1 and figure 1). Figure 1 is really confusing as it indicates almost all outcomes will be evaluated 4 times (6, 12, 24, 36 months after randomisation), including thyroid function, which is inconsistent with table 1 and the texts in methods section. Given that all the outcomes have been presented in Figure 1, I wonder whether Figure 1 adds any useful information here.

Response: Thank you for your careful reviewing. We have revised the methods and table 1 to keep the content consistent. We want to show the process of trial clear through Fig 1.(Please see Page 12, Line 21-22 and Page 16, Line 3-4)

The revisionsare as follows:

8. The thyroid function of patients will be examined before HSG, and at 4, 8, 12, and 24 weeks, and 9–12 months after HSG. We will detect free triiodothyronine (FT3), free thyroxine (FT4), thyroid stimulating hormone (TSH), antithyroglobulin antibodies (TgAb), and antithyroid peroxidase antibodies (TPOAb) levels at different time-points [27,28]. Also, we will record the thyroid function of patients as possible as we can to make the data more accurate.

9. As for the thyroid function of the patients, we will use ANOVA to compare it at different time-points [27].

- "Pregnancy leading to live birth is defined as the ratio of live births to clinical pregnancies." The intention of using pregnancy leading to live birth is for the analysis of time-to-event outcomes, and not for the analysis of dichotomous outcomes. It is poor methodological practice to use the clinical pregnancy as the denominator for the analysis of live birth. I would suggest considering this outcome as a time-to-event outcome (time to pregnancy leading to live birth) and replace "time to pregnancy resulting in ongoing pregnancy" by "time to pregnancy leading to live birth". This also implies in the statistical analysis section. Please also clarify the time points at which all pregnancy outcomes are measured – only at 36 months or also at 6 months?

Response: Thank you for your suggestion. We have made some changes of the secondary outcomes and statistical analysis section. (Please see Page 13, Line 8-18 and Page 15, Line 1-22 and Page 16, Line 1-6)

The revisionsare as follows:

> The rates of live birth rate/ clinical pregnancy / miscarriages/ ectopic pregnancy will be assessed. Live birth is defined as the birth of at least one living child; clinical pregnancy is defined as an ultrasound visible gestational sac; miscarriage is defined as a spontaneous loss of pregnancy; ectopic pregnancy is defined as an embryo implanted outside the uterine cavity. Each occurrence of one of these events will be recorded during the three-year follow-up.
>
> The time to ongoing pregnancy is defined as the time from randomization to the first day of the last menstrual period plus 4 weeks. It will be considered when ongoing pregnancy occurs. The time to first live birth is defined as the time from randomization to the date of the firth live birth. It will be considered when live birth occurs.

2. Sample size calculation.

- The references used in the response letter to justify the sample size calculation included mostly women with PID, which is different from the trial population. The authors may wish to cite some data from their clinic or other papers on women with low risk of tubal disease to justify the sample size calculation.

Response: Thank you for this concern.We have found that the baseline (such as the age) and the total pregnancy rate in women with a low risk of tubal diseases in China are similar to that in the population in Dreyer's research. Therefore, we used the ongoing pregnancy rate in the WSCM group from Dreyer 2017 for the sample size calculation without further justification.(Reference: Jie M, Mingming L, Ying M, et al.Diagnose Value of Hysterosalpingography on Tubal Infertility. Clinical Imaging Technology 2018;33(7):69-72)

- The wording on lost to follow up is inconsistent with the way the authors calculated the data. "Additional 10%" would imply the total sample size =1.1* calculated sample size. The authors may wish to rephrase the last two sentences for sample size calculation (Page 14, Line 20-22) as "Thus, 1040 women (520 in each group) will need to be randomised, including 10% loss to follow-up or protocol variation."

Response: Thank you for the suggestion. We have revised our manuscript as your suggestion to make the sample size calculation clear.(Please see Page 14, Line 20-21)

3. Statistical analysis.

- The current statistical analysis is almost identical to the relevant paragraphs in Dreyer 2017, apart from the tense used. I would leave it to the editors to decide whether it is acceptable or not.

- As the authors do not have a separate statistical analysis plan, they need to make this section more comprehensive according to the SPIRIT checklist, including any planned subgroup analysis, sensitivity analysis and the handling of missing data. Apart from intention

to treat analysis, whether per protocol analysis is also planned.

Response: Thank you for this concern. We have revised the part of statistical analysis in the protocol.(Please see Page 15, Line 1-22 and Page 16, Line 1-13)

Some further detailed comments:

Page 1

Title: When asked to "incorporate the population into the title", the authors should consider the most important characteristic of the population. In this case, the ethnicity does not seem to make a difference here. I would suggest replacing "Chines women" by "women with a low risk of tubal disease" (or pathology or something similar) if word limit allows. Please also remove the first "contrast" in the title.

Response: Thank you for the suggestion. We have amended the title as "Effects of oil-soluble versus water-soluble contrast media at hysterosalpingography on pregnancy outcomes inwomen with a low risk of tubal disease: study protocol for a randomized controlled trial".(Please see Page 1, Line 1-4)

Page 2 Abstract:

Introduction: Please incorporate the control group the objective.

Methods: Please emphasize that the study population is women with a low risk of tubal pathology.

Outcomes: Please refer to comments above on outcome definitions.

Response: Thank you for your careful reviewing. We have revised the abstract as your suggestions.(Please see Page 2, Line6-10 and Page 2, Line 12-14 and Page 2, Line 19-22)

The revisionsare as follows:

As the largest multi-centre randomized controlled trial (H2Oil trial from the Netherlands) has shown the oil-soluble contrast at HSG can enhance the fertility comparing to the water-soluble contrast, we propose this study to answer the question of whether the use of oil-soluble contrast media results in increased pregnancy rates in Chinese women undergo HSG.

The patients with a low risk of tubal disease will be randomized to undergo HSG using iodinated oil injection (OSCM group, oil-soluble contrast media) or ioversol injection (WSCM group, water-soluble contrast media).

The secondary outcomes will consist of thyroid function of patients, neonatal thyroid function, pain scores during HSG, live birth rate, clinical pregnancy rate, miscarriages, ectopic pregnancies, the time to ongoing pregnancy, the time to live birth, cost calculations of the OSCM group/WSCM group, and assisted reproductive technology (ART) treatments between two groups.

Page 3 Strengths:

Please do not overemphasize the "first" as this is neither the first study with long-term follow-up, nor the first study measuring pain. Long-term follow-up has been also performed in H2Oil trial. There have been many other trials reporting pain as outcome since 1990s, including the recent H2Oil Trial. Reporting long-term follow up itself is sufficient as a strength.

Response: We apologise for the improper expressions. We have revised the strengths.(Please see Page 3, Line 9-14)

Page 5, line 10: "did not pay close attention to" – did not report

Page 6, line 4-5: As lipiodol is not the product of interest, "lipiodol" – OSCM

Response: Thank you for your careful reviewing. We have revised the contents as your suggestions.(Please see Page 5, Line 7-8 and Page 6, Line 1-2)

The revisionsare as follows:

However, the review did not include Chinese population and did not report relevant contents about the thyroid function or long-term effects.

Previous studies showed that women with subclinical hypothyroidism were more prone to OSCM induced overt hypothyroidism [22], which may be due to the long half-life of OSCM excretion [23].

Page7, line 17: As post-wash total motile sperm count is only available for couples undergoing IUI, I don't know what "post-wash total motile sperm count" refer to here.

Response: We apologize for the confusion. We have amended the definition of the fertility of the partner.(Please see Page 7, Line 13-15 and Page 8, Line 12-13)

The revisionsare as follows:

1. Subfertility of at least one year and a fertile partner (defined as sperm count >15× 10^6 spermatozoa/mL or a post-wash total motile sperm count > 3 × 10^6 spermatozoa/mL before Intrauterine insemination (IUI)) [25].
2. Male subfertility defined as sperm count <15× 10^6 spermatozoa/mL or a post-wash total motile sperm count < 3 ×10^6 spermatozoa/mL before IUI.

Page 10, line 5: "Consecutive"- Eligible

Response: Thank you for your careful reviewing. We have revised the contents as your suggestions.(Please see Page 7, Line 1)

The revision is as follows:

Eligiblepatients will be asked to participate in this study after receiving oral and written information from a gynaecologist in outpatient services when tubal testing has been indicated and will be scheduled.

Page 15, line 22. "Monitoring staff will consist of an Independent Data Monitoring Committee (IDMC) and an ethics committee". The authors still did not describe the composition of IDMC. How many members? Who are they (if not established yet, at least mentioning the number of clinical experts and number of statisticians)?

Response: Thank you for this concern. We have completed the description of the composition of IDMC.(Please see Page 16, Line 17-18)

The revision is as follows:

The IDMC is composed of five members, including one statistician, three clinical experts and one ethicist.

Page 16, line 7. "An Independent Data Monitoring Committee (IDMC) and ethics committee will review data annually during the accrual period ….. At each meeting, the IDMC will be asked to give advice on whether the accumulated data from the trial, ……" As the trial is only planned to recruit patients for 12 months, it seems inappropriate to use "annually" "at each meeting" here.

Response: We are sorry for the carelessness. We have revised the content in theprotocol.(Please see Page 17, Line 2)

The revision is as follows:

An Independent Data Monitoring Committee (IDMC) and ethics committee will review data trimonthlyduring the accrual period and near the time that is planned for interim analyses.

Page 18, line 8. Please update the trial status if applicable.

Response: We have updated the trial status in the protocol.(Please see Page 24, Line 12-14)We postponed the start of the recruitment to prepare the work for the trial more carefully.

The revision is as follows:

Trial status

The recruitment starts from August 1stand we expect to complete the recruitment in August 2021.