PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Vaginal progesterone as luteal phase support in natural cycle frozen-thawed embryo transfer (ProFET): protocol for a multicenter, open-label, randomized controlled trial
AUTHORS	Stadelmann, Caroline; Bergh, Christina; Brännström, Mats; Olsen, Kristbjörg Heiður; Khatibi, Ali; Kitlinski, Margareta; Liffner, Susanne; Lundborg, Eva; Rodriguez-Wallberg, Kenny; Strandell, Annika; Westlander, Göran; Widlund, Gabriella; Magnusson, Åsa

VERSION 1 – REVIEW

REVIEWER	Li, Wentao Monash University, Department of Obstetrics and Gynecology
REVIEW RETURNED	16-Apr-2022

GENERAL COMMENTS	A three-arm trial protocol for understanding the effect of progesterone supplement in natural cycle FET. The sample size looks ample, and the design looks mostly appropriate. Some clarifications are likely further improve the transparency of the trial report. 1. Line 209-211. A blood example that measures progesterone will be taken before the intervention and the results will not be disclosed to patients. Does this imply the sample analysis is real-time and clinicians will know the results soon? If so, this may introduce performance bias, e.g., clinicians may think a patient allocated to the intervention group has too high serum progesterone to take exogenous progesterone, causing protocol violation. 2. The randomization will be stratified, but will the randomization list be based on permuted blocks? If yes, fixed or variable size? 3. Blinding. Any plan to mask embryologists and research personnel who perform patient follow up? 4. Outcome measurements. This section would be improved by adding definitions. Plus, progesterone day 3 takes place before the intervention, this is not an outcome. 5. Sample size. Line 290. It is not clear what 'a difference between groups of 8%' used for all comparisons above means. It looks all right for 3 weeks of progesterone (0.41) versus control (0.33), but 600 per group will not be sufficient for 7 weeks (0.49) vs 3 weeks (0.41), which requires at least 631 per group. So please specify the risk difference for each pairwise comparison. 6. Line 293. What variables were meant for imputation here? Outcomes and/or covariates? And what methods will be used for imputation? 7. Statistical analysis. As the randomization is stratified, the unadjusted analysis is unnecessary and may lead to wrong

	conclusions due to too liberal CIs. Line 310, please specify 'other predefined important predictors'. 8. It is ambitious to recruit 1800 participants in less than 2.5 years. Helpful if some justification regarding the feasibility is provided, such as the capacity of centers and the current speed of recruitment. 9. One minor issue. Line 121. I suppose the authors mean 'has almost eliminated the risk of OHSS'.

REVIEWER	Ng, Ernest The University of Hong Kong, Department of Obstetrics and Gynecology
REVIEW RETURNED	14-May-2022

GENERAL COMMENTS	The aim of this study protocol is to investigate if progesterone as luteal phase support increases live birth rate compared with no progesterone and if 7 weeks of progesterone treatment is more effective than the 3 weeks of treatment. Strength of the study 1. A multi-centre study 2. Natural cycles timed with LH surge 3. Transfer of one blastocyst after thawing Major comments 1. Introduction: The reasons of luteal phase insufficiency in stimulated IVF cycles should be stated as this gives the rationale of giving luteal phase support. However, the rationale in frozen embryo transfer in natural cycles is lacking. As the secondary objective is to compare the duration of luteal phase support, the controversy of the duration of luteal phase support should be summarized. This gives the readers a better background to understand the study protocol. On the other hand, the elective freezing can be shortened to improve the readability of the paper. 2. Treatment and intervention: L209-210 Blood will be taken in the morning 3 days after the LH-surge, before any start of progesterone. Serum progesterone level is not measured in the mid-luteal phase and certainly not after the start of vaginal progesterone. How can serum progesterone level at this time point be correlated with the outcome? Why is serum progesterone level not checked on the day of blastocyst transfer as it is mid-luteal phase in the control group and after progesterone in the treatment group (3 and 7 weeks of progesterone)? 3. Page 32 of 39 in the protocol: 10.4 Sample size calculation Groups are expected to have different live birth rates. For example: 3 weeks of progesterone will have 38% and 41% live birth rates while 7 weeks have 41% and 46%. The authors need to explain this. Minor comments 1. Abstract: L78 A typo, should be NCT04725864 2. Study design: L183 enrollment began in May 2021 and plan to continue till June 2024 according to the information stated in trial registry. The authors should clarify and align the dates.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1:
A three-arm trial protocol for understanding the effect of progesterone supplement in natural cycle FET. The sample size looks ample, and the design looks mostly appropriate. Some clarifications are likely further improve the transparency of the trial report.

1. Line 209-211. A blood example that measures progesterone will be taken before the intervention and the results will not be disclosed to patients. Does this imply the sample analysis is real-time and clinicians will know the results soon? If so, this may introduce performance bias, e.g., clinicians may think a patient allocated to the intervention group has too high serum progesterone to take exogenous progesterone, causing protocol violation.
Answer: The blood sample is taken after randomization. The results will be concealed and thus not affect group allocation. Clinicians performing embryo transfer will not know the result of S-progesterone on the day of embryo transfer, nor will the clinicians prescribing progesterone to the patient. We have clarified this in the manuscript on line 225: "The result will not be available to the patient, neither to the treating clinician until the end of the study."

2. The randomization will be stratified, but will the randomization list be based on permuted blocks? If yes, fixed or variable size?
Answer: No, the randomization is not based on block randomization. The randomization is based on stratification. The following stratification variables are used:
Previous ET not resulting in positive pregnancy test, number (0-2, ≥3)
Parity 0/>1
Age (<35/≥35) years
Treatment site 1-9 (now 1-13)

3. Blinding. Any plan to mask embryologists and research personnel who perform patient follow up?
Answer: The embryologists at participating clinics are not involved in the trial other than selecting the best embryo for embryo transfer. Statistician performing the analyses will be blinded.

4. Outcome measurements. This section would be improved by adding definitions. Plus, progesterone day 3 takes place before the intervention, this is not an outcome.
Answer: Thank you for bringing this to our attention. We have added definitions as a supplemental file, to avoid a longer manuscript, see the newly added supplemental file. We certainly agree that S-progesterone is not an outcome. We will study the association between progesterone levels day LH+3 and IVF outcome. Particularly, in patients with low progesterone levels we will compare outcome between patients receiving and not receiving progesterone support.

5. Sample size. Line 290. It is not clear what 'a difference between groups of 8%' used for all comparisons above means. It looks all right for 3 weeks of progesterone (0.41) versus control (0.33), but 600 per group will not be sufficient for 7 weeks (0.49) vs 3 weeks (0.41), which requires at least 631 per group. So please specify the risk difference for each pairwise comparison.
Answer: Thank you for pointing out that a clarification is needed. We have made the following clarification, seen as highlighted text in the revised manuscript, lines 302-326.

Sample size calculation
"In order to find an effect size of a 7% increase in LBR per randomized patient, measured as a difference in proportions between no progesterone (0.33) and any progesterone group (0.40), 1800 subjects are needed if allocated 1:2. In order to find a difference between no progesterone (0.33) and progesterone for 3 weeks (0.41) and 7 weeks (0.41) respectively, 1200 subjects are needed if

allocated 1:1. Also, for the comparison between the progesterone groups, 1200 subjects are needed if allocated 1:1, to detect a difference of 8%, (0.38 for 3 weeks of progesterone vs 0.46 for 7 weeks. For all comparisons above, except for the primary analysis, a difference between groups of 8% is used. If 1800 women are allocated 1:1:1, 600 to no progesterone, 600 to progesterone 3 weeks and 600 to progesterone 7 week, all four sample size calculations are fulfilled under the condition of a power of 0.80, a significance level 0.05 and a two-sided Fisher's exact test.

We thus have two primary superiority analyses in this study. The first is the comparison of LBR between no progesterone and the combined group of any progesterone with Fisher's exact test on significance level 0.05. If this test is significant the probability mass of 5% will be transferred to the second comparison of live birth between progesterone for 3 weeks compared with progesterone for 7 weeks. If the first test is significant, we have been able to show that any progesterone gives significantly higher LBR than in women without progesterone. If in the second comparison 7 weeks shows significantly higher LBR than 3 weeks, we have also confirmed superiority regarding 7 weeks over 3 weeks. If the first analysis is non-significant, we have not been able to show any confirmative results in this study. The comparisons between no progesterone and 3 weeks progesterone and between no progesterone and 7 weeks progesterone is performed to calculate mean difference with 95% CI between these groups."

We have changed "LBR per FET " to "LBR per randomized patient" in the manuscript since the sample size is calculated in this way. See line 77 and 303.

6. Line 293. What variables were meant for imputation here? Outcomes and/or covariates? And what methods will be used for imputation?
Answer: The manuscript has been changed to the following text:

Statistical analyses
"The main analyses will be on the full analysis set (FAS) without imputation. Complementary analyses will be performed on the per protocol population. The primary efficacy analysis regarding live birth will be conducted with multivariable logistic regression adjusting for all stratification variables on the FAS population. The first sensitivity primary analysis will be the same analysis also adjusted for the following other predefined important predictors:

- body-mass index
- smoking status
- duration of subfertility
- previous miscarriage (yes/no)
- blastocyst (day 5/day 6 at cryopreservation)
- number of embryos transferred

The second sensitivity primary analysis will be the same analysis as the primary efficacy analysis, but performed on the intention-to-treat population with multiple imputation of 100 datasets. Both primary outcome and stratified variables will be imputed. For adjusted analyses between two groups, multivariable logistic regression will be used for dichotomous variables. If model assumption is fulfilled the corresponding model with link=log will be given to present adjusted relative risk (RR) with 95% CI. For continuous variables analysis of covariance (ANCOVA) will be used for adjusted analysis between two groups.
Explorative unadjusted mean difference between the two groups with 95% CI will be given for dichotomous variables and continuous variables together with effect sizes. For continuous variables these 95% CI will be based on T-test or Fisher's non-parametric permutation test. For dichotomous variables relative risk and odds ratio will be given with 95% CI. Proportions will be given with exact 95% CI.

For comparison between two groups Mantel – Haenszel Chi square test will be used for ordered categorical variables and Fisher's exact test for dichotomous variables.
The distribution of continuous variables will be given as mean, standard deviation (SD), median, first and third quartiles (Q1, Q3), minimum and maximum. All significance tests will be two-sided and conducted at the 5% significance level."

7. Statistical analysis. As the randomization is stratified, the unadjusted analysis is unnecessary and may lead to wrong conclusions due to too liberal CIs. Line 310, please specify 'other predefined important predictors'.
Answer: Other predefined important predictors for live birth are:

- body-mass index
- smoking status
- duration of subfertility
- previous miscarriage (yes/no)
- blastocyst (day 5/day 6 at cryopreservation)
number of embryos transferred

8. It is ambitious to recruit 1800 participants in less than 2.5 years. Helpful if some justification regarding the feasibility is provided, such as the capacity of centers and the current speed of recruitment.
Answer: It certainly is a challenge to recruit a large number of patients in a relatively short time. The recruitment plan is based on the assumption that 50% of eligible patients accept participation in the trial. We have recruited 506 patients and randomized 420 patients by todays date (8th of June 2022) one year after the first clinic started recruitment. Most clinics started in the autumn 2021, and two will start in the autumn 2022. However, to increase recruitment-rate we recently identified four new centers in Sweden who joined the trial in April 2022. The total number of performed FET-NC in the year of 2020 at the twelve Swedish centers, which now are involved in the ProFET trial, were 3479 cycles.

9. One minor issue. Line 121. I suppose the authors mean 'has almost eliminated the risk of OHSS'.
Answer: Yes, thank you for pointing out this incorrect wording. We have now changed line 123 to the correct word order. See highlighted text:
"The freeze-all concept is also now widely used when pending risk of OHSS, and has almost eliminated the risk of OHSS, a potentially life-threatening condition(12-14)."

Reviewer #2:
Comments to the Author:
The aim of this study protocol is to investigate if progesterone as luteal
phase support increases live birth rate compared with no progesterone and
if 7 weeks of progesterone treatment is more effective than the 3 weeks of treatment.
Strength of the study
1. A multi-centre study
2. Natural cycles timed with LH surge
3. Transfer of one blastocyst after thawing

Major comments
1. Introduction: The reasons of luteal phase insufficiency in stimulated IVF cycles should be stated as this gives the rationale of giving luteal phase support. However, the rationale in frozen embryo

transfer in natural cycles is lacking. As the secondary objective is to compare the duration of luteal phase support, the controversy of the duration of luteal phase support should be summarized. This gives the readers a better background to understand the study protocol. On the other hand, the elective freezing can be shortened to improve the readability of the paper.
Answer: Thank you for your feedback. We have updated the manuscript based on your comments, see highlighted text in the revised manuscript, lines 119-156:

“Recently, the freeze-all concept has gained high popularity all over the world. Several large trials, comparing freeze-all vs fresh transfer, have shown similar live birth rates in ovulatory patients(7-10) while freeze-all has been shown to be beneficial in anovulatory patients(11). The freeze-all concept is also widely used when pending risk of ovarian hyperstimulation syndrome (OHSS), and has almost eliminated the risk of OHSS, a potentially life-threatening condition(12-14).

The most efficient protocol for FET is still not known. A Cochrane review, including 18 RCTs, comparing different cycle regimens for FET, comprising a total of 3815 women did not support one treatment modality over another when investigating LBR, however, with low certainty of evidence(15). Safety aspects in ART are of great importance in treatment decision. Recently, interest has risen concerning the role of the corpus luteum (CL) in frozen cycles and studies evaluating the risks of altered vascular adaptation associated with pregnancies following FET according to the presence or absence of CL have been published(16). The CL, developing after ovulation, is known to produce estrogen and progesterone, but also relaxin, a hormone that regulates the maternal cardiovascular and renal systems and hence mediates the hemodynamic changes occurring during pregnancy. In a prospective cohort study including almost 700 women, programmed cycles (artificial cycles using estrogen and progesterone for endometrial preparation) in FET with no CL present were associated with an almost three-fold increased risk of preeclampsia compared with modified natural cycles (natural cycles triggered by human chorionic gonadotrophin (hCG), for ovulation) with one CL present(16). Furthermore, in a recent Swedish large registry study, including almost 10 000 pregnancies/deliveries after FET, doubled rates of both hypertensive disorders of pregnancy and postpartum hemorrhage were found in programmed cycles compared to natural cycles(17). These studies thus support FET in natural cycles.

Luteal phase support (LPS) in fresh IVF cycles has been proven mandatory(18). Less is known regarding the role of LPS with progesterone in natural FET cycles. A natural ovulatory cycle would suggest that no supplementation needs to be given. However, the luteinizing hormone peak – used as a urine sample to detect ovulation - does not guarantee a subsequent ovulation. Furthermore, several studies have shown that corpus luteum deficiency with midluteal serum progesterone levels <10ng/ml could be a reason to support implantation and early pregnancy with LPS, even in a cycle where ovulation has occurred(19, 20). A study from 2018(21) showed that low but also high levels of progesterone were associated with a reduction in clinical pregnancy rate and LBR compared to normal levels. This has also been confirmed in a more recent study(22). Not only the doses, but also the duration of luteal phase support is widely discussed and differ between studies”.

2. Treatment and intervention: L209-210 Blood will be taken in the morning 3 days after the LH-surge, before any start of progesterone. Serum progesterone level is not measured in the mid-luteal phase and certainly not after the start of vaginal progesterone. How can serum progesterone level at this time point be correlated with the outcome? Why is serum progesterone level not checked on the day of blastocyst transfer as it is mid-luteal phase in the control group and after progesterone in the treatment group (3 and 7 weeks of progesterone)?
Answer: Please see answer to Reviewer 1, point 4. The idea behind blood samples being drawn before the intervention is of observational nature. We wish to investigate if there is an association between progesterone levels day LH+3 and IVF outcome, in the group not receiving any progesterone for luteal support. It may also be possible, in patients with low progesterone levels to compare

outcome between patients receiving and not receiving progesterone.

3. Page 32 of 39 in the protocol: 10.4 Sample size calculation Groups are expected to have different live birth rates. For example: 3 weeks of progesterone will have 38% and 41% live birth rates while 7 weeks have 41% and 46%. The authors need to explain this.
Answer: We have rephrased the text about the sample size calculation to avoid the misunderstanding that the same groups have different live birth rate estimations. Se Reviewer 1, point 5. (In manuscript lines 299-323).

Minor comments
1. Abstract: L78 A typo, should be NCT04725864
Answer: Thank you for noticing this. We have corrected the trial registration number to the right number.

2. Study design: L183 enrollment began in May 2021 and plan to continue till June 2024 according to the information stated in trial registry. The authors should clarify and align the dates.
Answer: The manuscript has been changed and is now corresponding to the dates in the trial registry, see line 194: “Patient enrollment began in May 2021 and is planned to continue until June 2024.”

VERSION 2 – REVIEW

REVIEWER	Li, Wentao Monash University, Department of Obstetrics and Gynecology
REVIEW RETURNED	23-Jun-2022

GENERAL COMMENTS	My comments have been appropriately addressed. Look forward to the findings.

REVIEWER	Ng, Ernest The University of Hong Kong, Department of Obstetrics and Gynecology
REVIEW RETURNED	22-Jun-2022

GENERAL COMMENTS	The authors had made revisions accordingly. I do not have further comments.