PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Acute Estradiol and Progesterone Therapy in Hospitalized Adults to Reduce COVID-19 Severity: A Randomized Control Trial
AUTHORS	Lovre, Dragana; Bateman, Kristin; Sherman, Mya; Fonseca, Vivian; Lefante, John; Mauvais-Jarvis, F

VERSION 1 – REVIEW

REVIEWER	Park, Jay The University of British Columbia
REVIEW RETURNED	04-Jul-2021

GENERAL COMMENTS	**Comments:** Lovre and colleagues aim to conduct a randomized clinical trial to evaluate the efficacy of estradiol (E2) and progesterone (P4) as a treatment option for hospitalized patients (WHO ordinal score of 3-5). This trial has already been reviewed by the US FDA in August 2020, which they have an IND for this experimental therapy. They have received their ethics approval for their study on May 14th, 2021. The proposed primary endpoint is the proportion of patients improving to scores 1 or 2 on the WHO scale through day 28, which will be analyzed using Pearson Chi-square test. The currently ongoing trial will aim to recruit up to 120 patients, and they will conduct an interim analysis at 50% enrollment. I believe this is of an interest to the audience of BMJ Open. There are some clarifications that should be made by the authors. The major comment I have for the authors is how they are going to recruit 120 patients in their trial. **Sample size calculation and choice in binary outcome** 1. The authors provide a table of required sample size (Table 2) with control event rates (CER) of 60% and 70% with treatment effect sizes in the absolute improvement to 90% and 95%. Even in the most conservative assumption, the authors are assuming 20% absolute improvement from 70% CER to 90%. This magnitude of treatment effects is likely unrealistic especially if we holistically view all of different treatments that have been evaluated for COVID-19. Could the authors provide their justifications to their assumptions in CER and desired effect sizes? Could the authors also justify why they decided to choose a binary outcome over time-to-event or ordinal analyses that likely will require fewer sample size? **Interim analysis** 2. The authors have stated that “[an] interim analysis will be performed at 50% enrollment.” (line 14 on page 15 of the pdf document). Could the authors confirm whether they

	mean the first and only interim analysis will be conducted when 60th patient has been enrolled into the trial or when this patient has finished their clinical follow-up? 3. Could the authors specify what CER and effect sizes they are assuming for this interim analysis at 60 patient and final recruitment of 120 patients. Assuming 70% CER and 20% absolute improvement in the treatment group, below is what level of interim analysis plans that should be reported, instead of just a single table they present. Sample size calculation for a binary endpoint Sequential analysis with a maximum of 2 looks (group sequential design). The sample size was calculated for a two-sample test for rates (two-sided), H0: pi(1) - pi(2) = 0, H1; treatment rate pi(1) = 0.9, control rate pi(2) = 0.7, power 80%. Stage 1 2 Information rate 50% 100% Efficacy boundary (z-value scale) 2.963 1.969 Overall power 0.1641 0.8000 Number of subjects 61.8 123.7 Cumulative alpha spent 0.0031 0.0500 Two-sided local significance level 0.0031 0.0490 Lower efficacy boundary (t) -0.377 -0.172 Upper efficacy boundary (t) 0.277 0.148 Legend: (t): approximate treatment effect scale **Recruitment** 4. One of the major challenges we have observed in COVID-19 trials has been recruitment. There are over 2,000 clinical trials that have registered for COVID-19 with the majority of the investigation occurring in the hospital setting. The authors here in the proposal manuscript have not specified any measures to recruit patients and minimize loss-to-follow-up. Could the authors specify how they are going to recruit patients from one academic center? According to one source, the 7-day average of COVID-19 cases in Louisiana has been 27 only. This is a great news for the people of Louisiana. It is likely that the investigators have already faced great challenges in recruitment. Could the authors comment on this? And how they are going to reach their 120 patient target?

REVIEWER	Stein, Donald Emory University
REVIEW RETURNED	13-Aug-2021

GENERAL COMMENTS	BMJ Open: bmjopen-2021-053684. Acute Estradiol and Progesterone Therapy in Hospitalized Adults to Reduce COVID-19 Severity: A Randomized Control Trial Precis: This manuscript reports on a proposed, FDA approved, clinically registered, pending study protocol for a Phase II trial testing both estradiol and progesterone in a Phase II, single center, randomized trial for treating the cytokine storm and subsequent pulmonary pathology caused by COVID-19. In this proposed trial patients will receive both estradiol cypionate and micronized progesterone along with standard of care (SOC)and be

	compared to a control group who will receive placebo and SOC. Both men and women patients will be included in the trial. Comments: • It would be helpful to provide the rationale for proposing a Phase II study rather than a Phase I, single center trial. • It would be helpful to provide the rationale for the trial design and provide more information on why the authors decided against using an adaptive trial design should there by signs of efficacy. • It would be important to note why the authors decided to combine both estrogen and P4 as a treatment when there is already evidence that P4 alone may be sufficient to alleviate COVID-induced lung pathology. • It would be helpful to provide more information about both the sources and forms of both estradiol and the progesterone along with the rationale for giving P4 orally and E2 by injection. The references and rationale for doing this would be very helpful to readers. One major concern is that the authors do not cite recent literature showing that P4 has already been used in a Phase I clinical trial for COVID-19 as has reported positive results. This is a glaring omission…especially since the study was recently conducted with FDA approval and published in the journal, CHEST. The specific reference: Progesterone in Addition to Standard of Care vs Standard of Care Alone in the Treatment of Men Hospitalized With Moderate to Severe COVID-19: A Randomized, Controlled Pilot Trial. Ghandehari S, et al Chest, 2021 Jul;160(1):74-84. doi: 10.1016/j.chest.2021.02.024. Epub 2021 Feb 20. PMID: 33621601 Free PMC article. Clinical Trial.

VERSION 1 – AUTHOR RESPONSE

REVIEWER: 1
Dr. Jay Park, The University of British Columbia

Lovre and colleagues aim to conduct a randomized clinical trial to evaluate the efficacy of estradiol (E 2) and progesterone (P4) as a treatment option for hospitalized patients (WHO ordinal score of 3- 5). This trial has already been reviewed by the US FDA in August 2020, which they have an IND for this experimental therapy. They have received their ethics approval for their study on May 14th, 2021. The proposed primary endpoint is the proportion of
patients improving to scores 1 or 2 on the WHO scale through day 28, which will be analyzed using P earson Chi-
square test. The currently ongoing trial will aim to recruit up to 120 patients, and they will conduct an i nterim analysis at 50% enrollment.

I believe this is of an interest to the audience of BMJ Open. There are some
clarifications that should be made by the authors. The major comment I have for the authors is how th ey are going to recruit 120 patients in their trial.
Response: As suggested by the reviewer we have addressed several points in the revised manuscript as stated below.

Sample size calculation and choice in binary outcome
1. The authors provide a table of required sample size (Table 2) with control event
rates (CER) of 60% and 70% with treatment effect sizes n the absolute improvement to 90% and 95% . Even in the most conservative assumption, the authors are assuming 20% absolute improvement fro

m 70% CER to 90%. This magnitude of treatment effects is likely unrealistic especially if we holisticall y view all of different treatments that have been evaluated for COVID-19.
Could the authors provide their justifications to their assumptions in CER and desired effect sizes? Could the authors also justify why they decided to choose a binary outcome over time-to-event or ordinal analyses that likely will require fewer sample size?

Response: Our sample size calculation was done in the second half of 2020 and was based on the percentage of subjects advancing to more severe disease (ICU transfer, intubation, death), which was estimated at 25% of hospitalized patients based on large published series (Richardson, 2020). The decision to use development of severe disease in our sample size calculation allowed calculating sample size accurately. The number of patients improved by WHO scale 1-3 has not been published at the time and would not have allowed us to calculate sample size since as no treatment was actually efficient to provide power calculation for a RCT. Furthermore, we will work closely with our statistician and conduct interim analysis at 50% of study completion to assess the number of subjects needed to reach statistical significance.
Although we chose binary outcome analysis overall, our primary and secondary outcomes combined will provide both the event and the timing of the event similar to what is normally seen in time-to-event outcomes.
Interim analysis
2. The authors have stated that "[an] interim analysis will be performed at 50% enrollment." (line 14 on page 15 of the pdf document). Could the authors confirm whether they mean the first and only interim analysis will be conducted when 60th patient has been enrolled into the trial or when this patient has finished their clinical follow-up?
Response: We are going to conduct the first and only interim analysis when the 60th subject has completed the end of the study visit.
3. Could the authors specify what CER and effect sizes they are assuming for this interim analysis at 60 patient and final recruitment of 120 patients. Assuming 70% CER and 20% absolute improvement in the treatment group, below is what level of interim analysis plans that should be reported, instead of just a single table they present.

1. Sample size calculation for a binary endpoint

Sequential analysis with a maximum of 2 looks (group sequential design).

The sample size was calculated for a two-sample test for rates (two-sided),

H0: pi(1) - pi(2) = 0, H1; treatment rate pi(1) = 0.9, control rate pi(2) = 0.7,

power 80%.

Stage	1	2
Information rate	50%	100%
Efficacy boundary (z-value scale)	2.963	1.969
Overall power	0.1641	0.8000
Number of subjects	61.8	123.7
Cumulative alpha spent	0.0031	0.0500
Two-sided local significance level	0.0031	0.0490
Lower efficacy boundary (t)	-0.377	-0.172
Upper efficacy boundary (t)	0.277	0.148

Legend:

Response: As suggested by the reviewer, we have included the table below in the revised manuscript.

Table 3. Interim analysis sample size calculation for a binary endpoint		
Stage	1	2

Information rate	50%	100%
Efficacy boundary (z-value scale)	2.963	1.969
Overall power	0.1641	0.8000
Number of subjects	61.8	123.7
Cumulative alpha spent	0.0031	0.0500
Two-sided local significance level	0.0031	0.0490
Lower efficacy boundary (t)	-0.377	-0.172
Upper efficacy boundary (t)	0.277	0.148
Sequential analysis with a maximum of 2 looks (group sequential design). The sample size was calculated for a two-sample test for rates (two-sided), H0: pi(1) - pi(2) = 0, H1; treatment rate pi(1) = 0.9, control rate pi(2) = 0.7. (t): approximate treatment effect scale		

Recruitment

4. One of the major challenges we have observed in COVID-19 trials has been recruitment. There are over 2,000 clinical trials that have registered for COVID-19 with the majority of the investigation occurring in the hospital setting. The authors here in the proposal manuscript have not specified any measures to recruit patients and minimize loss-to-follow-up. Could the authors specify how they are going to recruit patients from one academic center?

5. According to one source, the 7-day average of COVID-19 cases in Louisiana has been 27 only. This is a great news for the people of Louisiana. It is likely that the investigators have already faced great challenges in recruitment.

Could the authors comment on this? And how they are going to reach their 120 patient target?

Response: As suggested by the reviewer we have addressed several points in the revised manuscript as stated below.

Recruitment and retention are always challenging in RCTs and usual remedy is to increase in recruitment efforts and sites or change the inclusion/exclusion criteria. We will not be changing our inclusion/exclusion criteria as those details were closely evaluated by the by the FDA before they granted us the IND. We will increase recruitment efforts and sites if needed. To improve recruitment and retention of our subjects we have started providing compensation for participation in the trial and will follow up with our patient via phone call at day 60. We are fortunate to have an efficient team comprised of an internist, who is a Co-Investigator of this trial, who directly admits patients with COVID-19 at our hospital and a dedicated study coordinator who is only recruiting subjects for our trial. Our internist is able to approach patients as soon as they present in the emergency room or if they are transferred. Our recruitment will also improve as we now can tell our patients about the evidence that progesterone in addition to standard of care (vs standard of care alone) decreases hospital stay and oxygen requirement.[1] Other trials at our center are using drugs that have not proven efficacy. When we submitted this protocol, in May 2021, infection rates and numbers were very low however, since then Louisiana in on the 4th wave which is the worst one so far. In addition, infection transmission rates and testing positivity rates have been the highest so far, which is helping our enrollment. Louisiana has one of the lowest vaccination rates in the country therefore the infections will continue to persist. Lastly, vaccinated sick subjects who are hospitalized are not as severely ill (as compared to earlier variants) which makes them likely eligible for our study.

1. Ghandehari S, Matusov Y, Pepkowitz S, et al. Progesterone in Addition to Standard of Care vs Standard of Care Alone in the Treatment of Men Hospitalized With Moderate to Severe COVID-19: A Randomized, Controlled Pilot Trial. Chest. Jul 2021;160(1):74-84.

REVIEWER: 2

Dr. Donald Stein, Emory University

Comments to the Author:
BMJ Open: bmjopen-2021-053684. Acute Estradiol and Progesterone Therapy in Hospitalized Adults to Reduce COVID-19 Severity: A Randomized Control Trial
Precis: This manuscript reports on a proposed, FDA approved, clinically registered, pending study protocol for a Phase II trial testing both estradiol and progesterone in a Phase II, single center, randomized trial for treating the cytokine storm and subsequent pulmonary pathology caused by COVID-19. In this proposed trial patients will receive both estradiol cypionate and micronized progesterone along with standard of care (SOC)and be compared to a control group who will receive placebo and SOC. Both men and women patients will be included in the trial.

Response: As suggested by the reviewer we have addressed several points in the revised manuscript as stated below.

1. It would be helpful to provide the rationale for proposing a Phase II study rather than a Phase I, single center trial.
Response: We chose Phase II for two reasons: 1. The safety of the E2 and P4 is well known after decades of studies and 2. Our statistical calculation for the subject number requirement to reach significance was in line with a Phase II study.

2. It would be helpful to provide the rationale for the trial design and provide more information on why the authors decided against using an adaptive trial design should there by signs of efficacy.
Response: We chose against an adaptive trial design for 3 reasons: 1. More complex to execute clinically and statistically than traditional fixed design, 2. Need for more predictable allocation of funding and coordinator time and 3. Need to streamline FDA review timeline.

3. It would be important to note why the authors decided to combine both estrogen and P4 as a treatment when there is already evidence that P4 alone may be sufficient to alleviate COVID-induced lung pathology.
Response: We combine E2 and P4 based on data showing E2 and P4 have a different mechanism of action providing different anti-inflammatory and immunomodulatory actions, our hypothesis is that they will synergize. The rational paper was authored by the PI of the study Dr. Mauvais-Jarvis.[2]
2. Mauvais-Jarvis F, Klein SL, Levin ER. Estradiol, Progesterone, Immunomodulation, and COVID-19 Outcomes. Endocrinology. Sep 1 2020;161(9)doi:10.1210/endocr/bqaa127

4. It would be helpful to provide more information about both the sources and forms of both estradiol and the progesterone along with the rationale for giving P4 orally and E2 by injection. The references and rationale for doing this would be very helpful to readers.
One major concern is that the authors do not cite recent literature showing that P4 has already been used in a Phase I clinical trial for COVID-19 as has reported positive results. This is a glaring omission…especially since the study was recently conducted with FDA approval and published in the journal, CHEST
The specific reference: Progesterone in Addition to Standard of Care vs Standard of Care Alone in the Treatment of Men Hospitalized With Moderate to Severe COVID-19: A Randomized, Controlled Pilot Trial. Ghandehari S, et al Chest, 2021 Jul;160(1):74-84. doi: 10.1016/j.chest.2021.02.024. Epub 2021 Feb 20. PMID: 33621601 Free PMC article. Clinical Trial.

Response: We appreciate the reviewer comment on the CHEST article. As cited in our protocol, we were aware of two different trials, one studying estradiol and one studying progesterone at the time of our submission to BMJ. We were unaware of the CHEST article publication from July 2021 as we submitted our protocol on May 20, 2021. We are now even more encouraged about potential success of our trial and hypothesize there will be synergy of estradiol and progesterone.
We purchased E2 from Pfizer (NDC: 0009-0271-01) and P4 from VIRTUS (NDC: 69543-375-10).

We chose E2 injection once for two reasons: 1. Long duration of action which ensures adherence and 2. Injection bypasses first-pass liver metabolism and does not increase risk of clotting which is already a concern with COVID-19 patients.
We chose P4 pills for a few reasons: 1. 200mg dose via pills is used in millions of menopausal women and has known good tolerability and low side effect profile; 2. The daily P4 intramuscular injections would have been large volume and painful which would have potentially caused subjects to withdraw from the study; and 3. Subcutaneous P4 is only produced by the Institut Biochimique SA (IBSA, Lugano, Switzerland) and would have been slightly difficult to obtain delaying the start of our study.
We have in the revised manuscript to include CHEST article information and citation as well as information about the source and rationale for giving E2 injection and P4 pills.

VERSION 2 – REVIEW

REVIEWER	Park, Jay The University of British Columbia
REVIEW RETURNED	05-Oct-2021

GENERAL COMMENTS	The authors have responded to my inquiry about the choice in binary primary endpoint (instead of an ordinal primary endpoint) and justification of the absolute treatment effects of 20% as the target effect size with the following: "Furthermore, we will work closely with our statistician and conduct interim analysis at 50% of study completion to assess the number of subjects needed to reach statistical significance. Although we chose binary outcome analysis overall, our primary and secondary outcomes combined will provide both the event and the timing of the event similar to what is normally seen in time-to-event outcomes." Do the authors mean they will calculate conditional power or predictive power to re-assess the sample size at the interim analysis? I disagree with the authors' justifications provided, and there are type II error concerns associated with choosing a binary outcome as their primary endpoint.

REVIEWER	Stein, Donald Emory University
REVIEW RETURNED	19-Sep-2021

GENERAL COMMENTS	The authors have made a good faith attempt to respond appropriately to the concerns I expressed in their initial submission. Given the current literature and the rationale for the trial that they provide, I think this report would be of substantial interest to clinicians and researchers engaged in seeking to ameliorate the disease pathologies caused by COVID 19.

VERSION 2 – AUTHOR RESPONSE

REVIEWER: 1

Dr. Jay Park, The University of British Columbia

The authors have responded to my inquiry about the choice in binary primary endpoint (instead of an ordinal primary endpoint) and justification of the absolute treatment effects of 20% as the target effect size with the following:

"Furthermore, we will work closely with our statistician and conduct interim analysis at 50% of study completion to assess the number of subjects needed to reach statistical significance. Although we chose binary outcome analysis overall, our primary and secondary outcomes combined will provide both the event and the timing of the event similar to what is normally seen in time-to-event outcomes."

Do the authors mean they will calculate conditional power or predictive power to re-assess the sample size at the interim analysis?

I disagree with the authors' justifications provided, and there are type II error concerns associated with choosing a binary outcome as their primary endpoint.

Response: We appreciate the reviewer's suggestion on using ordinal outcomes as opposed to binary measures to assess efficacy of treatment.

Our primary focus is on improvement from a hospitalized category (WHO, 3-5) to a less severe ambulatory category (WHO 1-2). Analyzing change in all possible ordinal WHO scores, ranging from -5 (from a pre-treatment 5 to a post treatment 0) to 7 (pre-treatment 1 to post treatment 8) has limitations, in that a -1 change from category 5 to a severe category 4 would be given the same weight as a -1 change from a category 3 to moderate category 2. Instead, we considered three possible post treatment ordinal categories of response, improving to category1-2; remaining in category 3-5; worsening to category 6-8, and recalculated sample sizes based on the Wilcoxon-Mann-Whitney test and an O'Brien-Castelloe approximation, with 80% power and a 5% significance level. Results are presented in the table for the effect size in our manuscript. Required sample sizes are close to our projected 120 subjects, depending on assumptions made regarding the percent staying in categories 3-5 or worsening to categories 6-8.

Assumed Conditional Probabilities

TREATMENT	PRE (3-5) POST (1-2)	PRE (3-5) POST (3-5)	PRE (3-5) POST (6-8)	Sample Size
SOC	70%	30%	0%	70

SOC+E2+P4	90%	0%	10%	70
SOC	70%	30%	0%	58
SOC+E2+P4	90%	5%	5%	58
SOC	70%	15%	15%	49
SOC+E2+P4	90%	5%	5%	49
SOC	70%	30%	0%	49
SOC+E2+P4	90%	10%	0%	49
SOC	70%	0%	30%	49
SOC+E2+P4	90%	0%	10%	49
SOC	70%	15%	15%	43
SOC+E2+P4	90%	10%	0%	43
SOC	70%	0%	30%	37
SOC+E2+P4	90%	10%	0%	37

O'Brien, R. G. and Castelloe, J. (2007), "Sample-Size Analysis for Traditional Hypothesis Testing: Concepts and Issues," in Pharmaceutical Statistics Using SAS: A Practical Guide, ed. A. Dmitrienko, C. Chuang-Stein, and R. D'Agostino, Cary, NC: SAS Institute Inc., Chapter 10, 237–271.

We are unable to change our statistical analysis plan at this time given that the study has already started (we have enrolled 10 subjects) and that the FDA approval for IND was obtained based on our current plan.

VERSION 3 – REVIEW

REVIEWER	Park, Jay The University of British Columbia
REVIEW RETURNED	30-Oct-2021

GENERAL COMMENTS	I appreciate the authors presenting their re-calculated sample sizes. However, it is not surprising that assuming a highly effective treatment effect (20% absolute improvement from 70% to 90%) shows similar sample size requirements whether a binary or an ordinal outcome is used as the primary endpoint. As the authors have noted, changing the SAP and protocol is obviously not ideal, but based on my experience, it's not an impossible task. I recognize there are likely funding and other practical challenges where the recruitment for this treatment cannot exceed the proposed target of 120 participants. I don't see how the authors think such a large magnitude of treatment effect is a realistic target for their trial or any other trials. But since there is nothing "wrong" with how the authors have designed their trial. I wish them luck with their trial moving forward.