PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A pilot double-blind randomised placebo-controlled dose-response trial assessing the effects of melatonin on infertility treatment (MIART): study protocol
AUTHORS	Fernando, Shavi; Osianlis, Tiki; Vollenhoven, Beverley; Wallace, Euan; Rombauts, Luk

VERSION 1 - REVIEW

REVIEWER	Marian Showell Auckland University Auckland New Zealand
REVIEW RETURNED	21-Jul-2014

GENERAL COMMENTS	Did the trialists consider live birth and adverse events as primary outcomes? line 16 pg 4 What were the constraints of the study? line 22 pg 7 Why are the trialists interested in patient sleepiness? line 28 page 10 Is there a wash out period, if so how long? line 21 pg 11 what will trialsits do about managing continuous data regarding imputation?

REVIEWER	Ben Mol Uni of Adelaide Australia My former institute has been paid for lectures and consulting for Ferring and Merck Serono
REVIEW RETURNED	23-Jul-2014

GENERAL COMMENTS	This manuscript outlines the plan for a randomized clinical trial studying the neurohormone melatonin as treatment oxidative stress for the outcome of in-vitro fertilization (IVF). Recent studies have reported that melatonin may protect against the oxidative stress of infertility treatment on gametes and embryos. The authors plan to recruit 160 women, of which only 40 in the placebo group, making it a combined proof-of-principle and dose findings study. One could imagine that the placebo group would be larger, or the study would compare placebo to the highest dose of melatonine only (two groups of 80). The introduction and rationale are well described.

	Methods The methods are well describe. I think the indications for IVF/ICSI should be provided and reported. Formally, I would say the patient in this study is the infertile couples, and the treatments are IVF_melatonin or OVF only. The primary endpoint could be ongoing pregnancy. Live birth should be reported. Why are 2. Currently enrolled in another interventional clinical trial 3. Concurrent use of other adjuvant therapies (eg. Chinese herbs, acupuncture) 6. Autoimmune disorders 7. Undergoing preimplantation genetic diagnosis (PGD) exclusion criteria? Why do the authors opaque enveloped when this is a placebo controlled trial? The cross-over design in infertility studies is strongly debated, and I think there is agreement now that such designs are biased. In view of the highly experimental character, I think the use of melatonin in subsequent cycles is inappropriate. The authors could consider continue the study medication over multiple cycles, or new randomization after the 1st cycle, but the cross-over designs has no scientific value. Khan KS1, Daya S, Collins JA, Walter SD. Empirical evidence of bias in infertility research: overestimation of treatment effect in crossover trials using pregnancy as the outcome measure. Fertil Steril. 1996 May;65(5):939-45. The sample size paragraph should in my opinion have a hypothesis that can be tested. In summary, this is a very important and timely trial. The protocol, however, could in my opinion benefit from some adjustments.

VERSION 1 – AUTHOR RESPONSE

Did the trialists consider live birth and adverse events as primary outcomes?
As this is a pilot study aimed at designing a larger single dose RCT with appropriate sample size, we have elected to choose the clinical pregnancy rate as the primary outcome. While we recognise the importance of live birth as an outcome measure in infertility research, there is also evidence that conclusions on the effectiveness of a treatment based on either clinical pregnancy or live birth as endpoints are comparable.1 Conclusions on clinical pregnancy rate are therefore reliable enough and are available on much shorter timeframes to allow us to proceed with the larger RCT which of course will have live birth rate as a primary outcome. Nevertheless, as it is compulsory in Australia to report all pregnancy outcomes to the Australian and New Zealand Assisted Reproduction Database (ANZARD) database, we will of course continue to track and report on these pregnancies as they develop.
Similarly, we anticipate a low rate of adverse outcomes, and therefore, this is also considered as a secondary outcome for this pilot study. As this is a pilot study, it has been recommended by others to

identify and define a single primary outcome in order to focus and maintain the integrity of the analysis .2

line 16 pg 4 What were the constraints of the study?
The constraints referred to here, include challenges that are expected in recruiting infertile patients for a study that includes a placebo arm. The experience in Australia with IVF-related RCTs has been that a control arm reduces participation as the study treatment is often already available in routine practice, as it is with melatonin. While this is an anticipated challenge, there is no way to be certain that it will significantly impact uptake until recruitment begins. However, our study design which guarantees at least one exposure to melatonin is expected to improve the recruitment.

line 22 pg 7 Why are the trialists interested in patient sleepiness?
Sleepiness, while being studied in different populations in relation to treatment with melatonin, has never been assessed in a dose-response fashion in the infertile population. We hope to use this information to further establish the safety and acceptability of melatonin treatment.

line 28 page 10 Is there a wash out period, if so how long?
The half-life of melatonin is 4-6 hours. Therefore, melatonin is completely removed from the system shortly after 24 hours from administration. Supplementation ends on the day of oocyte collection (approximately 5 days before embryo transfer). Once patients have had their first stimulated cycle of treatment, a second stimulated cycle will routinely not commence until at least 4 weeks after the negative pregnancy test (approx. 6 weeks after last melatonin dose), much longer than the elimination time for melatonin. In our opinion this is more than adequate as a washout period.

line 21 pg 11 what will trialists do about managing continuous data regarding imputation?
The SPSS multiple imputation routine (MCMC algorithm known as Fully Conditional Specification) will be used to handle missing continuous data.

One could imagine that the placebo group would be larger, or the study would compare placebo to the highest dose of melatonin only (two groups of 80).
While we understand that this would help improve the power of this study for a single dose analysis, our clearly stated intention is to perform a pilot study primarily concerned with finding an optimal dose and estimated effect size in order to inform an adequate sample size for a larger single dose RCT. As such, to limit the analysis to a single dose would defeat our purpose. Similarly, if a dose-effect can be demonstrated, this would assist in proving causality. 3

I think the indications for IVF/ICSI should be provided and reported
We plan to both record and report this. We apologise for not including this in the study protocol article. It has now been included.

Formally, I would say the patient in this study is the infertile couples, and the treatments are IVF_melatonin or OVF only. The primary endpoint could be ongoing pregnancy.
We thank the reviewer for highlighting these points.
We have addressed the issue of ongoing pregnancy being the primary endpoint (see above) and provided a reference to support our argument. 1 This reference has also been added to the manuscript in the 'outcome' section.
In terms of defining the treatment unit, we agree that the infertile couple is usually the treatment unit. However, donor sperm is not an exclusion criteria and as such not all women will be undergoing treatment as a couple. We are not convinced that this distinction makes a difference in the analysis of the data.

Live birth should be reported.

We will report this as a secondary outcome (please refer to Table 1 under 'Clinical outcomes'). Please see above for justification of why this is a secondary and not a primary outcome.

Why are the following exclusion criteria?
2. Currently enrolled in another interventional clinical trial
If a participant is also currently enrolled in another trial this may have the effect of contaminating the sample, for example, as concomitant trials may be using other medications and interventions. Secondly, our ethics committee does not permit patients to be concurrently recruited for 2 separate trials.
3. Concurrent use of other adjuvant therapies (eg. Chinese herbs, acupuncture)
If a participant is using other adjuvant therapies, it will become difficult to determine to what extent any measured effect is due to melatonin alone. Our aim with this study is to maintain a dataset that is as free of potential confounding factors as possible. Patients in our trial will be starting their first cycle and adjuvant use in the "therapy-naïve" population is usually minimal anyway.
6. Autoimmune disorders
While melatonin has a benign safety profile, case reports have suggested a role in the exacerbation of rheumatoid arthritis, autoimmune hepatitis and multiple sclerosis. While these far from prove causality, we feel that it is unethical to treat patients with autoimmune disorders with melatonin as part of this trial.
7. Undergoing preimplantation genetic diagnosis (PGD)
Again, inclusion of this group of patients may result in 'contamination' of the dataset. If some patients undergo PGD and others do not, differences in the primary outcome may not be due to the study intervention.

Why do the authors opaque enveloped when this is a placebo controlled trial?
Although we accept that the placebo nature of the trial should limit the risk of unblinding investigators or patients, association of particular treatment effects (e.g. sleepiness) with certain randomisation codes may nevertheless become apparent during the trial and to prevent further allocation bias proper concealment of treatment allocation is necessary.

The cross-over design in infertility studies is strongly debated, and I think there is agreement now that such designs are biased. In view of the highly experimental character, I think the use of melatonin in subsequent cycles is inappropriate. The authors could consider continue the study medication over multiple cycles, or new randomization after the 1st cycle, but the cross-over designs has no scientific value.
We would like to make two points.
Firstly, we agree that the cross-over design is strongly debated, but many, including the venerable Douglas Altman, still see several benefits for certain types of studies. 4 In particular, as alluded to above, patient recruitment is known to be significantly enhanced if patients know they will be guaranteed access to the treatment arm. In our experience, this is a major issue with parallel arm RCTs that study interventions that are already accessible to IVF patients (which is also true for melatonin). In addition, the ability to carry out within and between subject comparisons is a particularly attractive feature of this design, assuming of course appropriate statistical modelling is used. To overcome the major concern of carry-over or residual treatment effect, although we have explained above that this risk is minimal, we intend to perform a sub-analysis on the data from the first cycle separately. This will also deal with differing drop-out rates in the first and second cycle.
Again we would like to emphasise that this is a pilot dose-finding study. We believe that the possible draw-backs are definitely outweighed by the benefits of our design. These include improved recruitment, larger sample size, and the possibility of within patient comparisons (with appropriate statistical techniques). Despite the lack of good evidence, melatonin use as an adjuvant is quickly becoming routine treatment. We intend to either provide preliminary evidence that this may be safe and effective practice or show that the effect is at best minimal. Our pilot data may then provide the

basis for a larger RCT to provide more conclusive proof.

The sample size paragraph should in my opinion have a hypothesis that can be tested.
We and other statisticians have argued that pilot trials should not aim to prove a testable alternative hypothesis but rather measure the effect size and its confidence intervals. But we agree that a testable hypothesis can be formulated for a pilot study without necessarily providing a sample size calculation. We have therefore added the hypothesis that we expect to see a dose-response relationship between melatonin administration and clinical pregnancy.

We are grateful for such a thorough review of our protocol article and for such a timely response from the journal. We hope that we have sufficiently addressed the queries and issues raised by the reviewers. We look forward to hearing from you in the near future.