ANSWERING REVIEWERS

Dear Editor,

Please find enclosed the edited manuscript in word format (file name: 19659-Review.doc).

Title: Towards better meta-analyses in assisted reproductive technology: Fixed, random or multivariate models?

Author: Philippe Lehert

Name of Journal: *World Journal of Meta-Analysis*

ESPS Manuscript NO: 19659

The manuscript has been improved according to the suggestions of reviewers:

1 Format has been updated.

2 Revision has been made according to the suggestions of the reviewer:

I am very grateful to the reviewer for having addressed these comments. To facilitate reading, the questions and comments are in italic, and they are directly followed by my comments.

(1) The author uses a mathematical approach to demonstrate existing concerns with meta-analyses conducted with studies on ART. In general, the author applies the findings from the analyses to support existing recommendations for "best practices" when performing a meta-analysis. The concern about selecting the appropriate model used for the meta-analysis based on heterogeneity of the studies included in the analysis is not new. Strategies proposed to address this concern and those regarding the stability of the findings in a meta-analysis include performing a sensitivity analysis, use a FM and RM in the meta-analysis, and provide a "level of confidence" for the finding based on the risk for bias from the meta-analysis.

I agree with the reviewer that the question of Random versus fixed is not new. However, my concern was an in depth discussion in the particular field of ART/IVF, where many meta-analyses are conducted and very often, the results are subject to controversy. Reviewer for the European Medicine Agency EMA, and Food and drug administration FDA, I observed that more and more meta-analyses are published with the simplest technique, although I suspected that this technique might probably provide strongly biased results in many cases. Besides, reviewing/authoring meta-analyses convinced me on considerable viewpoint divergences, highly debated and controversial in particular between statisticians and ART clinicians. This motivated me to undertake this research, for helping ART experts to designate the best meta-analyses and inviting reviewers to accept modern standards.

From the observed results, another important finding is that – at least in ART- using a sensitivity analysis by using FM and RM in the meta-analysis, and provide a level of confidence for the finding based on the risk for bias from the meta-analysis, is perhaps not useful, as the random model covers the whole question, and in particular the Q test is a "false friend" in that it doesn't reject homogeneity assumption in most of the cases

I am convinced that my results have a very important practical implication for ART/IVF researchers: I provide a practical example based on IVF data easily interpretable for clinicians; on the other hand, my conclusions are very simple and strictly specific for meta-analyses in this pathology.

(2) As the author noted, multivariate meta-analysis makes presumptions about the selected studies and need additional information (see appendix). The findings from the analyses conducted in the manuscript are most relevant to ART. Would this article attract more readers in a journal on reproductive medicine than in this one?

Thank you for this relevant comment. Of course, I did this work mainly for ART pathology, as I saw so many meta-analyses in this field with this invariably chosen fixed model, that I decided to publish a specific question on it. However, I think this paper might be (modestly, of course..) a seed for a new way at looking at meta-analyses: In every pathology, the rcandomized controlled Trials have specific conditions, that might considerably impact the relevance of particular models of meta-analysis. Just as an example, in Cardiology, I have the experience that study effect is very often of a limited importance, which may help to recommend a particular protocol of meta-analysis for all the systematic reviews in this pathology. In ART, where the center effect has by nature a huge effect, it was highly expecyed that the random model should be used by principle. Thus, in conclusion, although my conclusions will be specific to ART meta-analyses, this paper might be of some help for researcher tempting a guideline approach pathology-specific.

(3) Please review use of language for clarity. For example, "comparing the FM with two alternative models" in the abstract and "general specific profile" in the materials (para one). Also consider stating that the FM was compared with the RM, and the univariate model (RM) was compared with the multi-variate model (MM) in the abstract because comparisons were limited to these variables (see para 4 in the discussion).

Dear reviewer, thank you for this request. You are correct in saying that the RM was compared with the MM. I completely reviewed my abstract in this purpose. For facility, please find the reviewed text (Accounting for limitations in words):

Aim

comparing the validity of the fixed, random, and multivariate meta-analytical models applied in meta-analyses in Artificial Reproduction Technique (ART).

Methods

Based on common characteristics of IVF meta-analyses, we simulated a large number of data to compare results issued from the Fixed Model (FM) with the Random Model (RM). For multiple endpoints MA, we compared the univariate RM with the multivariate model (MM). Finally, we illustrate our findings in re-analyzing a recent meta-analysis.

Results

In our review, although a homogeneous effect was excluded in 89% of the MAs (11%), FM was utilized in 41 studies (82%). From simulations, a concordance of 59±6% was found between the two tests, with up to 65% of falsely significant results with FM. The Q-test on studies characterized by substantial heterogeneity falsely accepted homogeneity in 46% of studies. Comparing separate univariate RM and MM on multiple endpoints studies, MM reduces the Between Endpoint Discrepancy (BED) of 68%, and increases the power of 57±8%. In the example dealing with the controversial effect of LH supplementation to FSH during ovarian stimulation in IVF cycles, MM reduced BED by 66%, and consistent effects were found for all the endpoints, irrespective of partial reporting.

Conclusion
The fixed model generally may produce falsely significant differences. The random model should always be used. For multiple endpoints, the multivariate model constitutes the best option.

(4) How was the number of simulations determined for this analysis?
I apologize for this lack of precision, however, this paper looked quite technical and I have tried to minimize details on computation. In this particular problem, the simulation aimed at a multivariate purpose, thus it was not possible to rely in a main statistics in particular. an infinite number of replications is theoretically requested. The key to the usefulness of the bootstrap is that it converges in terms of numbers of replications reasonably quickly, and so running a finite number of replications is good enough—assuming the number of replications chosen is large enough. Under this principle: I have chosen a first guess in programming 2,500 replications. I changed the random-number seed and checked the difference which provided a overall relative error of 7% which was too much. I restarted with 20,000, and through iteration, I observed a convergence at around 9700 replications, thus I repeated the whole analysis in using n=10000. This took a time of a whole week end of computer elapsed time.

To comply with the comment of the reviewer, I changed accordingly in the text, following this (see statistical Analysis section):

"…The parameters of simulation were based on the distribution of the following variables: NST, number of reported endpoints, level of partial reporting, between study heterogeneity, effect size quantified by the Risk Ratio, within-study correlations between effects. The replication number was fixed through an exploratory research in assessing the convergence of the estimator so that the relative change does not exceed .01. 10,000 simulated samples were needed for this, and were generated from a multivariate distribution of these parameters accounting for their correlation. The Results distributed according to a normal distribution are reported as mean±SD, and otherwise by median and interquartile range IR…"

3 References and typesetting were corrected.

Thank you again for publishing our manuscript in the *World Journal of Meta-Analysis*.