Author's response to reviews

Title: Using Bayesian statistics to estimate the likelihood a new trial will demonstrate the efficacy of a new treatment

Authors:

David Biau (david.biau@aphp.fr)

Samuel Boulezaz (samuel.boulezaz@gmail.com)

Laurent Casabianca (laurent.casabianca@aphp.fr)

Moussa Hamadouche (moussa.hamadouche@aphp.fr)

Philippe Anract (philippe.anract@aphp.fr)

Sylvie Chevret (sylvie.chevret@aphp.fr)

Version: 1 **Date:** 31 May 2017

Author's response to reviews:

Answers to reviewer's comments.

Reviewer 1

1. My major concern is the method section. The authors didn't clearly present the model and method used for Bayesian posterior probability and predictive probability. Neither did they provide sufficient reference to justify the specific method they used.

Answer: We appreciate the reviewer's comments and we have improved the method section according to his/her remarks, as detailed below.

1.1 page 8, line 13-30. Need to provide reference for the Bayesian beta-binomial meta-analysis model. Is fixed effector or random effect model used? What's the rational for your choice?

Answer: Standard metaanalyses are often based on fixed or random effects models. In the fixed-effects model, it is assumed that the parameter of interest is identical across studies and the difference between the observed proportion and the mean is only due to sampling error. In the random-effects model, the observed difference between the proportions and the mean cannot be entirely attributed to sampling error but may rely to many other unobserved factors. By contrast, we used a Bayesian approach that considers the parameter of interest to be a random variable, thus, incorporating the potential difference between the proportions as the random-effects model.

The main interest of the Bayesian approach in this setting was to further allow probabilistic statements, and that was the main rational of our choice.

This has been more clearly stated and a reference for the Bayesian beta-binomial model has been added in the revised manuscript.

1.2 Page 8, line 35. What's the reason for 5% and 10% risk difference? Is it choosed by clinical judgment?

Answer: The cutoffs of 5% or 10% difference were indeed chosen on the basis of clinical judgment. These risk differences were chosen on a clinical basis. They represent what surgeons would call a small clinical effect and a reasonable clinical effect.

This has been reported in the revised manuscript.

1.3 Page 8, line 39-42. What's the method for predictive probability? And, one key component for predictive probability is the decision rule at the end of trial. Namely, how to decide the experimental treatment is in favour? By hypothesis testing (e.g., at 5% significance level)? Or, by posterior probability (e.g., posterior probability of risk difference at least 5% >80%)?

Answer: The predictive probabilities were computed as previously reported, based on the posterior distribution of the difference in response rates, and considering the sample size of the new trial to be enrolled. ("We then computed the predictive probabilities that given the results of all previous published trials (a priori information) the next scheduled trial would achieve a risk difference of observed failure rates of at least 5% or 10% in favor of the experimental treatment.") More precisely, we used a MCMC algorithm to provide these computations. This has been more clearly mentioned in the revised manuscript.

1.4 Page 8, line 42-46. The whole sentence is not readable. For example, 95% coverage probability.

Answer: We agree with the reviewer that the sentence was not readable. It has been rephrased in the revised manuscript.

1.5 What's the method and calculation for the third question? I'm not able to find it in the introduction.

Answer: We based the likelihood that a new planned trial shifts the overall evidence accumulated in the literature on simulated trials. Indeed, we simulated samples of patients with response rate in the control arm drawn from the last posterior (obtained at the end of the metaanalysis), with varying sample sizes and failure probability in the intervention arm, then computing the posterior probability of reaching a difference of at least 5% between arms. We agree with the reviewer and apologize to have not mentioned this method in the manuscript. This has been described in the revised manuscript.

1.6 Page 8, line 53. Need to provide reference for 'DerSimonian and Laird method'.

The following reference has been added: DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7:177–188.

Answer: This has been done

2. To help reader understand the application of this manuscript, it's necessary to introduce the disease and the experimental/conventional treatment.

Answer: A brief description of the pathology and treatment has been added in the method section.

"For instance, a current question among the orthopaedic community is whether patient-specific instrumentation, a recent innovative technology used during total knee replacement to improve implant positioning, is superior to conventional instrumentation [9]. To ensure the long term success of a knee replacement it is paramount that the best limb alignment (180°) is achieved during the operation: knees which deviate by more than 3 degrees from this angle are more likely to fail early. The standard procedure to ensure limb alignment is to use intra-medullary jigs. Recently, patient specific guides based on a preoperative scanner or MRI have been developed to improve the precision of limb alignment during the surgery. Seventeen studies and 10 reviews or meta-analyses have been published to address the issue in less than four years and there is still no evidence for a difference between both treatments."

3. Other line-by-line comments.

3.1 Page 1, 'As a worked example, we conducted a Bayesian cumulative meta-analysis to summarise' should be 'summarize'

Answer: This has been corrected accordingly

3.2 Page 5, line 55-59. 'We therefore conducted a Bayesian cumulative meta-analysis of patient-specific compared to conventional instrumentation in patients undergoing total knee replacement to achieve adequate alignment.' Need to re-write this sentence. It's not readable.

Answer: The sentence has been rewritten and simplified. We therefore conducted a Bayesian cumulative meta-analysis of patient-specific instrumentation compared to conventional instrumentation in patients undergoing total knee replacement.

3.3 Page 9, line 1-2. Need to provide reference for R2jags and rmeta packages.

Answer: The following references have been added:

https://cran.r-project.org/web/packages/R2jags/index.html

https://cran.r-project.org/web/packages/rmeta/index.html

3.4 Page 9, line 42. '(95% CrI: 20.7; 27.7%%)'. Remove the extra '%'.

Answer: This has been corrected accordingly

3.5 Page 10, line 1-2. What's the conclusion for the first research question?

Answer: Actually, the answer was that the likelihood that the experimental treatment is superior to the control treatment is marginal at best. This has been reported in the revised manuscript.

What is the likelihood that the experimental treatment is superior to the control treatment given the evidence accumulated so far?

The posterior probabilities that the proportion of failures in the experimental group be inferior to the proportion of failures in the control group by 5% and 10% was less than 5% after trial #4 and trial #2 respectively. After all the available evidence, these probabilities were 4.2% and 0.013% (Table 2). The likelihood that the experimental treatment is superior to the control treatment is therefore marginal at best.

3.6 Page 10, line 26-27. Is it calculated by simulation? What's the simulation setting?

Answer: See answer above to point 1.5 that details the simulated trials.

Reviewer 2

In the manuscript, the authors conducted a Bayesian cumulative meta-analysis to summarize the benefit of patient-specific instrumentation on the alignment of total knee replacement from previously published evidence, using a cumulative meta-analysis approach. Overall, the proposed research is interesting and can be useful in meta-analysis, in particular in addressing the question whether the experimental treatment is superior to the control treatment on the basis of previous evidence available. To improve the manuscript, I have a few comments and suggestions:

1) In page 6 (see the Methods section), the authors claimed that they examined the studies up to 01 January 2016. And in Table 1, the most recent study included was published on 04 March 2015. I understand that the manuscript was probably prepared in early 2016, but given that it is now in the middle of 2017, to update the studies as well as the manuscript seem to be a must for the paper being considered for publication.

Answer: We do appreciate the Reviewer's concern regarding the publications used in this study. However we would like to raise a few points against doing so. First, since our last review and the publication of Molicnik's paper in March, 2015, only one relevant randomized controlled trial has been published (Huijbregts HJ Component alignment and clinical outcome following total knee arthroplasty: a randomised controlled trial comparing an intramedullary alignment system with patient-specific instrumentation. Bone Joint J. 2016 Aug;98-B(8):1043-9.). Nevertheless, the main focus of our paper was not to prove or disprove the clinical hypothesis, but rather to develop the method and illustrate it using a real data example. Moreover, based on the data accumulated so far in 2015, the main result of our paper is that the addition of any other trial is of limited value because we show that the probability that the experimental treatment be superior to the control treatment was less than 5% after 13 trials.

2) In page 7, the authors proposed a Bayesian analysis to estimate the posterior probabilities and their posterior credible intervals. It is unclear to me if the proposed algorithm was referred from the existing literature (if so, please provide some necessary references in the manuscript). If, instead, the proposed algorithm is new, then I would expect that many practitioners will not be able to follow the algorithm in the Appendix and apply the proposed method in their own studies. From this point of view, I strongly encourage the authors to create a software or R package and make it freely available to the public.

Answer: Actually, the Bayesian analysis of binomial variables is not new, and based on simple algorithms (when dealing with failure probabilities in each arm), and simple MCMC algorithms (when computing difference in these probabilities across arms). The Appendix details the approach so that one could apply it easily using standard statistical software.

3) I have some serious concerns on the statistical analysis, in particular in the claim of the following sentence in page 7: "We used the I2 test to estimate heterogeneity across trials, with P<0.1 being considered significant". First of all, the "I2" is not used for the purpose a test (and hence will not produce a P value), but rather used to measure the heterogeneity of the included studies. For the testing purpose, one often employ the Q test or other methods. I would hence suggest the authors to find a statistician to double check the statistical analysis part (including the Appendix on Bayesian analysis) to make sure the statistical claims are all correct.

Answer: We agree with the reviewer's comment and we apologize for this mistake. It has been corrected in the revised manuscript:

I2 was used to quantify heterogeneity and we used the Q chi-squared statistic to test heterogeneity across trials with P<0.1 being considered significant (I2 = 45%, Q=0.04)

Some minor comments:

4) The title is relatively long and not very readable (hence a better title is desired).

Answer: We have shortened the title:

Using Bayesian statistics to estimate the likelihood a new trial will demonstrate the efficacy of a new treatment.

5) In page 7, I could not follow the sentence "Summary statistics were computed (median with interquartile range [IQR] or percentage". What does it mean?

Answer: We have removed this sentence since it is not critical.

6) The paper writing is not very careful, where I found quite a few typos and grammar mistakes. As one obvious example, in page 7 line 28 there is "… for each of each of …"

Answer: We have corrected this typo and others in the manuscript.

7) The Discussion section is somehow redundant. In particular, it is not needed to repeat the three questions on the "likelihood" in page 10 lines 8-15.

Answer: We have shortened the discussion and we made it more focused.