PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Recombinant human activated protein C for the treatment of severe sepsis and septic shock: a study protocol for incorporating observational evidence by using Bayesian approach
AUTHORS	Zhang, Zhongheng

VERSION 1 - REVIEW

REVIEWER	Adrian Barnett Queensland University of Technology, Australia
REVIEW RETURNED	15-May-2014

GENERAL COMMENTS	This protocol details a meta-analysis that will combine data from RCTs and observational data. The observational data will inform the prior of the meta-analysis of RCTs. This is an interesting approach. The paper may be too technical for BMJ Open readers and may be better suited to a statistics journal. Why does table 1 include the filter of "clinical trials" if a major focus of this paper is to include observational data? The exclusion criteria are not clear. Does "replicated cohort population" mean a study where the data has already been published? And does "include single arm" mean studies without a control sample? Will all observational studies will be included regardless of their quality? It's not stated how the RCT quality data will be used. Will sensitivity analyses be run based on only including studies of a certain quality? The planned investigation of publication bias is thorough. It's not clear how the hyper-parameter mean of 0.33 for the overall pooled effect was generated. The paper states it comes from the mean of the log odds ratio from the observational studies. How many observational studies were there? Can these studies and the key results be listed? How were these observational studies found, using a systematic review? I think this is important as the results of the meta-analysis could strongly depend on this mean. We therefore need to see as much detail on the observational studies as we do the RCTs. The WinBUGS code was useful, but it would be worthwhile pointing out that this is based on just three values of alpha, whereas the text

	uses 12. Perhaps it would be better to change the WinBUGS code to include the 12 alpha's. The analysis will give 12 results, one for each alpha weighting. How will these results be clearly interpreted? Will an attempt be made to identify the "best" weighting? I am concerned that presenting 12 results will confuse most clinicians. What guidance will be given to interpret the results? Minor comments - Page 2, Line 16 "using a Bayesian" - Page 2, Line 19, "trial" not "trail" (also on page 4) - Page 4, Line 6, "outcomes" not "outocmes" - Page 4, line 30, typo "Contrlled" The paper was generally clear, but a grammar check would be useful.

REVIEWER	Martí-Carvajal, Arturo Iberoamerican Cochrane Network, Venezuela. I am corresponding author of "Marti-Carvajal AJ, Sola I, Gluud C, Lathyris D, Cardona AF. Human recombinant protein C for severe sepsis and septic shock in adult and paediatric patients. The Cochrane database of systematic reviews. 2012;12:CD004388.
REVIEW RETURNED	25-May-2014

GENERAL COMMENTS	1. Drug was withdrawn in Octuber 2011 due to non-clinical benefit and high risk for bleeding. 2. A Cochrane review show a high quality of evidence for rejecting more studies on APC.

REVIEWER	Rujipat Samransamruajkit MD. Pediatric Critical Care Division Department f Pediatrics King CHulalongkorn Memorial Hospital Faculty of Medicine Chulalongkorn University
REVIEW RETURNED	25-May-2014

GENERAL COMMENTS	The authors should better clarify in more details is in the future trial we are using bayesian approach what the results would be the different than current data. Should discuss more on details if we want to use what are the reasons, and would happen need to give more examples of current negative study in critical care medicine such as using HFOV in adult, A protocolized adult septic shock RCT etc.

REVIEWER	Andre Kalil University of Nebraska Medical Center, Omaha, NE, USA
REVIEW RETURNED	28-May-2014

GENERAL COMMENTS	The authors proposed a new protocol to evaluate the effectiveness of activated protein C for the treatment of sepsis and septic shock through the use of Bayesian methodology. Comments: The use of “activated protein C” in the title is too broad and confusing. I recommend the use of more specific terms: either “recombinant human activated protein C” or “drotrecogin alfa activated”. This drug was not used or intended for use in patients with “sepsis” only. Please change to “severe sepsis” for a more appropriate title and literature search. The introduction states that only RCTs have been included in published meta-analysis on this therapy. However, this is incorrect since there is one meta-analysis that included observational studies (Lancet Infect Dis 2012;12:678-86). I suggest the authors to also take advantage of this published meta-analysis to compare their literature findings, as well as to make sure they don’t miss relevant observational and randomized studies on their new literature search. On page 5, data extraction, the authors need to be more specific: please included which mortality outcome will be collected and analyzed, i.e. 28-day? In-hospital? 60-day? 90-day? Also, explain how you will combine different outcome follow ups from different studies. Their data extraction should include disease severity scores such as APACHE II, SOFA, and SAPS for 2 reasons: 1) this therapy has shown effect modification based on disease severity and this needs to be incorporated into the Bayesian model, and 2) most RCTs and cohorts have reported severity scores. The authors mentioned they will collect odds ratios, but they need to know that the RCTs and some of the cohorts used risk ratios. Thus RR needs to be collected as well to avoid missing relevant information. In addition, the authors need to explain which outcome metrics they will chose for the Bayesian analysis, i.e. OR or RR, and how they will combine both metrics, that is, which transformation methods they will use for adequate comparability. For safety, authors state that they will extract only “major bleeding”, please be aware that you may miss other relevant bleeding events if you don’t perform the literature search (and safety analysis) by using other common terms such as “severe bleeding”, “acute bleeding”, “life-threatening bleeding”, “significant bleeding”, “cerebral bleeding”, “intracranial hemorrhage”, and “gastro-intestinal hemorrhage or bleeding”. The authors state that they will use the Delphi list to assess the quality of the RCTs, but they don’t state how they will evaluate the quality of the observational studies. I suggest them to use a validated method such as the Newcastle-Ottawa scale. A new results section called “sensitivity analysis” should be added to evaluate possible confounding factors, potential subgroups of

	interest (shock vs. non-shock), and interaction terms, such as rhAPC effect by disease severity. The statistical analysis section needs major revision: The methods section lacks technical details; the model needs to be specified more clearly; instead, they refer to the code in table 3, which has no comment – the reader should not have to read the authors' code to figure out what their model is. The authors need to present their specific model in statistical terms in the methods section. This also includes the code in table 3, which needs a more extensive explanation. Explicit rationale for the choice of priors needs to be clearly stated in the methods section. Also, parameter estimates, posterior densities, and joint probabilities for the posterior distributions need to be included in methods. Diagnostics for convergence (MCMC) need to be provided, as well as the demonstration that the treatment effects met the exchangeability assumption, and that the data followed a normal hyperdistribution. This protocol is lacking the entire results section. Please provide a detailed description of what you intend to include in and how you intend to present the Bayesian findings in the results section. Last, this protocol would benefit from a more in depth English language revision.

VERSION 1 – AUTHOR RESPONSE

Reviewer Name Adrian Barnett
Institution and Country Queensland University of Technology, Australia
Please state any competing interests or state 'None declared': None declared

This protocol details a meta-analysis that will combine data from RCTs and observational data. The observational data will inform the prior of the meta-analysis of RCTs. This is an interesting approach. The paper may be too technical for BMJ Open readers and may be better suited to a statistics journal.

Why does table 1 include the filter of "clinical trials" if a major focus of this paper is to include observational data?

The "filter" function was deleted in the revised version, to include both RCTs and observational trials.

The exclusion criteria are not clear. Does "replicated cohort population" mean a study where the data has already been published? And does "include single arm" mean studies without a control sample?

The exclusion criteria were revised according to the reviewer's comments as follows:" We will exclude studies that 1) do not report mortality as an endpoint; 2) are secondary analysis of a primary study that its data have been published in anther paper; 3) include single arm that no comparison is made

between different treatment strategies (e.g. such as analysis of risk factors)."
Hope this will address the reviewer's concern about its uncertainty.

Will all observational studies will be included regardless of their quality?

Observational studies are subject to bias and the sources of such bias are often not easy to be quantitatively analyzed. Therefore, we included observational evidence as a prior with different degrees of skepticism (alpha power) by using Bayesian approach. Therefore, the quality of observational studies was not included as an exclusion criterion. We will include all observational evidence for analysis. We explicitly stated it in the method section in the revised manuscript. Furthermore, the quality of observational studies will be assessed as described in table 3.

It's not stated how the RCT quality data will be used. Will sensitivity analyses be run based on only including studies of a certain quality?

We added sensitivity analysis by excluding studies with poor quality in methodological design in revision.

The planned investigation of publication bias is thorough.

It's not clear how the hyper-parameter mean of 0.33 for the overall pooled effect was generated. The paper states it comes from the mean of the log odds ratio from the observational studies. How many observational studies were there? Can these studies and the key results be listed? How were these observational studies found, using a systematic review? I think this is important as the results of the meta-analysis could strongly depend on this mean. We therefore need to see as much detail on the observational studies as we do the RCTs.

The mean of prior distribution (the figure 0.33 in the expression is used for illustration purpose, and is not obtained from real analysis) is the natural log of pooled OR (LOR) estimated from observational data. The pooled OR is estimated with Bayesian approach with random-effects model.
The observational evidence is obtained by using systematic review of the literature and the protocol has been defined in the manuscript, the observational studies to be included are: 1) cohort studies using multivariable analysis with aPC treatment as one of the covariates; 2) cohort studies using propensity analysis; 3) case-control studies; 4) both prospective and retrospective designs will be included. 5) all observational studies will be included irrespective of their quality in methodological design.
Because this is a study protocol, how many observational studies will be included is still unknown at this stage. We will present all the information after we completed this project, and the results will be published in full details.

The WinBUGS code was useful, but it would be worthwhile pointing out that this is based on just three values of alpha, whereas the text uses 12. Perhaps it would be better to change the WinBUGS code

to include the 12 alpha's.

We changed it to accommodate 12 alpha.

The analysis will give 12 results, one for each alpha weighting. How will these results be clearly interpreted? Will an attempt be made to identify the "best" weighting? I am concerned that presenting 12 results will confuse most clinicians. What guidance will be given to interpret the results?

Actually we acknowledge that there is no criterion to choose which evidence is the best one, we just want to show how different degrees of skepticism will change the RCT meta-analysis. One advantage is to quantitatively assess the skepticism by using Bayesian approach.

Minor comments
- Page 2, Line 16 "using a Bayesian"
- Page 2, Line 19, "trial" not "trail" (also on page 4)
- Page 4, Line 6, "outcomes" not "outocmes"
- Page 4, line 30, typo "Contrlled"

The paper was generally clear, but a grammar check would be useful.

Typo errors were corrected and grammar was checked in the revised manuscript.

Reviewer Name Martí-Carvajal, Arturo
Institution and Country Iberoamerican Cochrane Network, Venezuela.
Please state any competing interests or state ‘None declared’: I am corresponding author of "Marti-Carvajal AJ, Sola I, Gluud C, Lathyris D, Cardona AF. Human recombinant protein C for severe sepsis and septic shock in adult and paediatric patients. The Cochrane database of systematic reviews. 2012;12:CD004388.

1. Drug was withdrawn in October 2011 due to non-clinical benefit and high risk for bleeding.

2. A Cochrane review show a high quality of evidence for rejecting more studies on APC.

Thanks you very much for letting me to comment on.

We agree with the reviewer that the best evidence has been provided by incorporating evidence from RCTs and the drug has been withdrawn from the market based on this evidence. However, our study will serve as a supplement to the study by the reviewer by showing how observational evidence will change the result obtained by RCT meta-analysis. Although RCT is still the golden standard for clinical decision making, it is not without limitations. Randomized controlled trial (RCT) is designed to test the biological efficacy of a certain treatment, whereas the observational study is to test the effectiveness of that treatment in real world setting. The differences between efficacy and effectiveness may result from issues related to trial design, patient selection, and therapeutic implementation. RCT is criticized for its non-real world setting, strict inclusion/exclusion criteria.

Reviewer Name Rujipat Samransamruajkit MD.

Institution and Country Pediatric Critical Care Division
Department f Pediatrics
King CHulalongkorn Memorial Hospital
Faculty of Medicine
Chulalongkorn University

Please state any competing interests or state ‘None declared’: None

The authors should better clarify in more details is in the future trial we are using bayesian approach what the results would be the different than current data.

The present study aims to investigate the use of Bayesian approach as a tool of model building to incorporating observational evidence into RCT meta-analysis. The question raised by the reviewer on what will the result of trials be changed is not the scope of the current study. Furthermore, this sentence is a little confusing in my sense, probably I might not catch up with its meaning.

Should discuss more on details if we want to use what are the reasons, and would happen need to give more examples of current negative study in critical care medicine such as using HFOV in adult, A protocolized adult septic shock RCT etc.

Clinical researches in CCM are full of negative studies and we added a section to discuss this phenomenon according to the reviewer’s comments.
Many RCTs in critically ill patients showed neutral effect of the intervention under investigation. In other situations the initial trials showed beneficial effect of the intervention which however was refused by subsequent meta-trial. Reasons for these negative results include timing of enrollment, endpoint selection and heterogeneous subjects.

Reviewer Name Andre Kalil
Institution and Country University of Nebraska Medical Center, Omaha, NE, USA
Please state any competing interests or state ‘None declared’: None declared.

The authors proposed a new protocol to evaluate the effectiveness of activated protein C for the treatment of sepsis and septic shock through the use of Bayesian methodology.

Comments:

The use of “activated protein C” in the title is too broad and confusing. I recommend the use of more specific terms: either “recombinant human activated protein C” or “drotrecogin alfa activated”.

I changed this term according to the reviewer’s comments.

This drug was not used or intended for use in patients with “sepsis” only. Please change to “severe sepsis” for a more appropriate title and literature search.

I changed this term according to the reviewer’s comments.

The introduction states that only RCTs have been included in published meta-analysis on this therapy. However, this is incorrect since there is one meta-analysis that included observational studies (Lancet Infect Dis 2012;12:678-86). I suggest the authors to also take advantage of this published meta-analysis to compare their literature findings, as well as to make sure they don't miss relevant observational and randomized studies on their new literature search.

Yes, we have referenced this meta-analysis in our study (introduction section). And we will take advantage of this published meta-analysis to compare their literature findings, as well as to make sure they don't miss relevant observational and randomized studies on their new literature search.

On page 5, data extraction, the authors need to be more specific: please included which mortality outcome will be collected and analyzed, i.e. 28-day? In-hospital? 60-day? 90-day? Also, explain how you will combine different outcome follow ups from different studies.

We strongly agree with the reviewer that the definition of mortality is an important contribution of the heterogeneity. Therefore, we specified how to handle the mortality in the manuscript:
Mortality is defined variably across studies (e.g. 28-day, In-hospital, 60-day or 90-day) and we will include all types of definitions for analysis. If there are sufficient number of studies with uniform definition (n>5), subgroup analysis will be performed.

Their data extraction should include disease severity scores such as APACHE II, SOFA, and SAPS for 2 reasons: 1) this therapy has shown effect modification based on disease severity and this needs to be incorporated into the Bayesian model, and 2) most RCTs and cohorts have reported severity scores.

We revised manuscript according to the reviewer's comments.

The authors mentioned they will collect odds ratios, but they need to know that the RCTs and some of the cohorts used risk ratios. Thus RR needs to be collected as well to avoid missing relevant information. In addition, the authors need to explain which outcome metrics they will chose for the Bayesian analysis, i.e. OR or RR, and how they will combine both metrics, that is, which transformation methods they will use for adequate comparability.

We used OR for calculation in Bayesian analysis, this has been described in the method section. About cohort studies reporting RR, we used standard formula to transform it to OR for further analysis (1/odds = (1-risk)/risk = (1/risk) − 1). The reference for this transformation has been cited in the manuscript (J Gen Intern Med. May 2008; 23(5): 635–640.).

For safety, authors state that they will extract only "major bleeding", please be aware that you may miss other relevant bleeding events if you don't perform the literature search (and safety analysis) by using other common terms such as "severe bleeding", "acute bleeding", "life-threatening bleeding", "significant bleeding", "cerebral bleeding", "intracranial hemorrhage", and "gastro-intestinal hemorrhage or bleeding".

To accommodate all these terms described by the reviewer, we rephrased this sentence by using the word "bleeding" and "hemarrhage" instead of the "major bleeding". Hope this will address the reviewer's concern.

The authors state that they will use the Delphi list to assess the quality of the RCTs, but they don't state how they will evaluate the quality of the observational studies. I suggest them to use a validated method such as the.

We agree with the reviewer that quality assessment for observational studies is of paramount important. And we added the tool Newcastle-Ottawa scale for its analysis. This tool is described in table 3 in revised manuscript.

A new results section called "sensitivity analysis" should be added to evaluate possible confounding factors, potential subgroups of interest (shock vs. non-shock), and interaction terms, such as rhAPC effect by disease severity.

A new section was added as follows:
Sensitivity or subgroup analysis
Sensitivity analysis will be performed by excluding studies with poor quality in methodological design. Subgroup analysis will be performed to explore the confounding factors such as shock versus non-shock, effect of aPC modified by disease severity. If there are sufficient number of studies with uniform definition of mortality (n>5), subgroup analysis will be performed by different mortality definitions.

The statistical analysis section needs major revision: The methods section lacks technical details; the model needs to be specified more clearly; instead, they refer to the code in table 3, which has no comment – the reader should not have to read the authors' code to figure out what their model is. The authors need to present their specific model in statistical terms in the methods section. This also includes the code in table 3, which needs a more extensive explanation.

We extensively revised the manuscript and the description of the model was divided into three parts: the first part describes how the model was adapted for current use (this has been described in another paper and we simply repeat it here for the ease of reading). The second part, after model development, we explained how to interpret the model. The third part is the interpretation of the code presented in table 4:

The framework to incorporate observational data as informative prior is presented by Chen and Ibrahim. The model development has been described elsewhere and we repeat it here for the ease of reading. Let the data from RCTs be denoted by D, and the likelihood of RCTs be denoted by $L(\theta|D)$. Suppose we have data from observational studies which is denoted by D0. Furthermore, let $P(\theta)$ denote the prior distribution for θ before observational studies are incorporated. $P(\theta)$ is the initial prior distribution for θ. Given α, the power prior distribution of θ is defined as
$P(\theta|D0, \alpha) \propto L(\theta|D0)\alpha \times P(\theta|c0)$
where c0 is the hyperparameter for initial prior, and α is used to weight observational evidence relative the likelihood of RCT evidence. The value of α controls the impact of observational evidence on $P(\theta|D0, \alpha)$. When evidence from RCTs is added to the model, a power transformation of the observational data likelihood is considered:
$P(\theta|Data)=L(\theta|RCTs)\times[L(\theta|Obs)]\alpha \times P(\theta)$
Where $P(\theta|Data)$ is the posterior distribution for model quantities, $[L(\theta|Obs)]$ is the likelihood function derived from observational data, and $L(\theta|RCTs)$ is the likelihood function from RCT data. The weight

of observational data is counted by the power α. The power takes values from 0 to 1. If α=0 the observational data are essentially removed from analysis and only RCTs are used for evidence synthesis; if α=1 observational data are taken at its “face value” and is not discounted at all. Traditional meta-analyses such as those done in Cochrane collaboration included only RCTs actually render α=0. In our analysis, alpha will take 12 values ranging between 0 and 1 (0.000001, 0.001, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0), resulting in a series of posterior distributions for OR. As shown in table 4, the WinBUGS code is composed of three parts: part (1) is to duplicate meta-analysis of RCTs for 12 times, once for each value of α to discount the observational evidence; part (2) is the meta-analysis model. In this section, i represents the component studies and k indices each of the twelve meta-analyses. These meta-analyses differ with each other only in the prior distribution for the overall pooled effect d, which is represented by the line:
d[k]~dnorm(0.33, prec.d[k])
The mean of prior distribution (the figure 0.33 in the expression is used for illustration purpose, and is not obtained from real analysis) is the natural log of pooled OR (LOR) estimated from observational data. The pooled OR is estimated with Bayesian approach with random-effects model. The code for the random-effects meta-analysis is shown in table 4. The precision of the prior distribution, prec.d[k], is determined in part (3). Part (3) is to calculate precision of prior discounted by using alpha.

Explicit rationale for the choice of priors needs to be clearly stated in the methods section. Also, parameter estimates, posterior densities, and joint probabilities for the posterior distributions need to be included in methods.

We acknowledge that the appropriate chose of alpha is a challenge because there is no rule on how observational evidence can be discounted. Therefore we explicitly chose a range of alpha values to see how RCT evidence can be influenced by different degrees of skepticism on observational evidence.

The parameter of OR was estimated. Posterior density will be expressed as OR and credible interval, the density graphs will not be displayed. The joint probability for posterior distribution will be reported (still in the form of OR and credible interval).

Diagnostics for convergence (MCMC) need to be provided, as well as the demonstration that the treatment effects met the exchangeability assumption, and that the data followed a normal hyperdistribution.

The following section has been added according to the reviewer’s comments:
Diagnostic for convergence will be explored by running two trains. Simulated values will be compared to identify when they become similar. History plots with different trains superimposed (in different colors) will help to determine convergence. Furthermore, we will use Brooks-Gelman-Rubin diagnostic to test convergence. The procedure will produce three colored lines (red, blue and green). Convergence is deemed to occur when the red line settles close to 1 and blue and green lines converge together.
The assumption of exchangeability and normal distribution were made before conducting Bayesian analysis. However, we do not know how to empirically prove these assumptions. This is just like performing sample size calculation that many assumptions can be made before a trial, but there is no way to scientifically prove them. If possible, could the reviewer provide some reference materials for us and we are extremely glad to learn more on this topic.

This protocol is lacking the entire results section. Please provide a detailed description of what you

intend to include in and how you intend to present the Bayesian findings in the results section.

Result section is conventionally missing for a study protocol. Only after completion of the analysis will the result be displayed.
However, we agree with the reviewer that explicit state of what to display will clarify the protocol, and I do so in the revised manuscript.

Last, this protocol would benefit from a more in depth English language revision.

English editing was made in the revised version.

VERSION 2 – REVIEW

REVIEWER	Andre Kalil University of Nebraska Medical Center, Omaha, NE, USA
REVIEW RETURNED	28-Jun-2014

GENERAL COMMENTS	I have 3 remaining comments on the authors' response to my first review: Authors: Yes, we have referenced this meta-analysis in our study (introduction section). And we will take advantage of this published meta-analysis to compare their literature findings, as well as to make sure they don't miss relevant observational and randomized studies on their new literature search. Reviewer's response: On page 22, the authors state the following about the paper by Kalil and LaRosa: "One meta-analysis has incorporated observational evidence at its face value, that is, effect sizes obtained from observational studies were treated as those obtained from RCTs.(10) The assumption of the equivalence of observational studies and RCTs is not valid." This statement is not correct. The paper aimed only to evaluate the effectiveness of rhAPC outside the RCTs, however the authors were asked by the reviewers to incorporate RCTs as part of a sensitivity analysis, i.e. to evaluate if the effectiveness seen with the observational studies remained similar after the incorporation of RCTs. Please remove your incorrect statement and replace it with "Kalil and LaRosa provided a frequentist analysis of both observational and randomized studies, but no Bayesian analyses were performed." Authors: We used OR for calculation in Bayesian analysis, this has been described in the method section. About cohort studies reporting RR, we used standard formula to transform it to OR for further analysis (1/odds = (1-risk)/risk = (1/risk) − 1). The reference for this transformation has been cited in the manuscript (J Gen Intern Med. May 2008; 23(5): 635–640.). Reviewer's response: This formula is not corrrect - this is the formula for calculating odds from risk, not OR from RR. The reference the authors cited provides a formula that converts OR to RR. They can derive the formula, but they need to assume a specific control event rate (CER). Authors: To accommodate all these terms described by the

	reviewer, we rephrased this sentence by using the word “bleeding” and “hemarrhage” instead of the “major bleeding”. Hope this will address the reviewer’s concern. Reviewer’s response: No, this does not address my concerns. The authors must separate major versus any bleeding, but to do that they will need to specify which combination of terms will be used to group the 'major bleeding', and which combination of terms will be used for 'any bleeding'.

VERSION 2 – AUTHOR RESPONSE

Reviewer Name Andre Kalil
Institution and Country University of Nebraska Medical Center, Omaha, NE, USA
Please state any competing interests or state ‘None declared’: None declared.

I have 3 remaining comments on the authors' response to my first review:

Authors: Yes, we have referenced this meta-analysis in our study (introduction section). And we will take advantage of this published meta-analysis to compare their literature findings, as well as to make sure they don’t miss relevant observational and randomized studies on their new literature search.

Reviewer’s response: On page 22, the authors state the following about the paper by Kalil and LaRosa: “One meta-analysis has incorporated observational evidence at its face value, that is, effect sizes obtained from observational studies were treated as those obtained from RCTs.(10) The assumption of the equivalence of observational studies and RCTs is not valid.” This statement is not correct. The paper aimed only to evaluate the effectiveness of rhAPC outside the RCTs, however the authors were asked by the reviewers to incorporate RCTs as part of a sensitivity analysis, i.e. to evaluate if the effectiveness seen with the observational studies remained similar after the incorporation of RCTs. Please remove your incorrect statement and replace it with “Kalil and LaRosa provided a frequentist analysis of both observational and randomized studies, but no Bayesian analyses were performed.”

We agree with the reviewer’s comments and revised the manuscript accordingly.

Authors: We used OR for calculation in Bayesian analysis, this has been described in the method section. About cohort studies reporting RR, we used standard formula to transform it to OR for further analysis (1/odds = (1-risk)/risk = (1/risk) − 1). The reference for this transformation has been cited in the manuscript (J Gen Intern Med. May 2008; 23(5): 635–640.).

Reviewer’s response: This formula is not corrrect - this is the formula for calculating odds from risk, not OR from RR. The reference the authors cited provides a formula that converts OR to RR. They can derive the formula, but they need to assume a specific control event rate (CER).

We derived the formula in the revised manuscript, and CER was indicated.

Authors: To accommodate all these terms described by the reviewer, we rephrased this sentence by

using the word “bleeding” and “hemarrhage” instead of the “major bleeding”. Hope this will address the reviewer’s concern.

Reviewer’s response: No, this does not address my concerns. The authors must separate major versus any bleeding, but to do that they will need to specify which combination of terms will be used to group the 'major bleeding', and which combination of terms will be used for 'any bleeding'.

We added the following sentence in the manuscript:
The adverse event of bleeding will be categorized into categories of major bleeding (terms consist the combinations of “massive”, “major” and “bleeding”, “hemorrhage”) and any bleeding (terms consist the combinations of “minor” and “bleeding”, “hemorrhage”).