PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Multivariable Fractional Polynomial Interaction to Investigate Continuous Effect Modifiers in a meta-analysis on higher versus lower PEEP for patients with ARDS
AUTHORS	Kasenda, Benjamin; Sauerbrei, Willi; Royston, Patrick; Mercat, Alain; Slutky, Arthur; Cook, Deborah; Guyatt, Gordon; Brochart, Laurant; Richard, Jean; Stewart, Thomas; Meade, Maureen; Briel, Matthias

VERSION 1 - REVIEW

REVIEWER	Diehl Service de Réanimation Médicale et INSERM UMR_S1140 Hôpital Européen Georges Pompidou. 20 rue Leblanc. Paris. France.
REVIEW RETURNED	09-Feb-2016

GENERAL COMMENTS	The authors present the results of a multivariable fractional polynomial interaction (MFPI) approach investigating the interactions between 4 continuous variables and the PEEP setting strategy, based on individual data from 3 major RCTs. They mainly found a non-linear effect modification of high PEEP by P/F ratio with reduced mortality for moderately severe ARDS patients; with a potent additional benefit of high PEEP in patients with high BMI. This is clearly a major topic from a clinical point of view. Obviously, the results extend our knowledge on this important topic, as compared to previous results published by the same group using pre-defined thresholds. However, as acknowledged by the authors, the results deserve further confirmations in external samples before modification of the current ventilator strategies in ARDS. Finally, the paper is well-written and I have very few comments. Major comment I have not sufficient knowledge of the statistical methodology for appreciating if the high percentage of missing values can affect the overall results, and if the imputation method is methodologically adequate. A statistical advice could be of value. Minor comment: Please check affiliation of Prof Laurent Brochard. Since the data are not really new, Figure 1 could perhaps be deleted. As driving pressure is known as the ventilation variable that best stratified the mortality risk, it could be interesting for the reader to know if it could be possible to include this important parameter in the MFPI approach.

REVIEWER	Elizabeth Colantuoni Bloomberg School of Public Health, Johns Hopkins University
REVIEW RETURNED	14-Mar-2016

GENERAL COMMENTS	The authors present an interesting re-analysis of an individual patient data meta-analysis of three randomized trials assessing the benefits of high vs. low PEEP. In meta-analysis, potential effect modifiers can start to be explored by pooling information across trials; however, it is often common to dichotomize continuous candidate effect modifiers to provide simplicity of interpretation. The authors discuss a recent approach where the treatment effect is estimated as a smooth function of the candidate continuous effect modifier within each study and then these functions are pooled across studies using weight functions defined for each study across the range of the candidate continuous effect modifiers. This paper would contribute the literature in two ways: a) this paper provides another example of how to implement and interpret the findings of a meta-analysis using fractional polynomials to estimate a treatment effect as a function of a continuous modifier, b) this paper contributes additional data beyond the original meta-analysis conducted on the three trials and offers additional insight into hypothesize effect modifiers for the effect of PEEP in patients with ARDS. My points below are places in the manuscript where I feel the authors can strengthen the applicability of the manuscript for ARDS researchers who may not be statisticians and also places where extra information or a clarification of language would improve the content of the manuscript. 1. Introduction: Second paragraph: The authors should better motivate/describe the Sauerbrei and Royston approach for the meta-analysis. The relevance of the discussion in this paragraph is not tied very well in to the meta-analytic approach. 2. Clinical outcomes: First paragraph: The authors should be more precise in defining the statistical approach and models for the time to event analyses. Specifically, the author state “Patients who died before achieving unassisted breathing within the first 28 days were censored at the day of death. With this procedure we circumvented the problem of competing risk in intensive care trials [22] in the analysis of this outcome.” However, this procedure does not circumvent the competing risk issue, it modifies the interpretation of the standard time-to-event analysis to a cause-specific model. This should be more precisely defined for those applicable time-to-event outcomes. 3. Clinical outcomes: Second paragraph: I’m not entirely clear why the considerations for the competing risk of death differ for the time to unassisted breathing and pneumothorax requiring chest tube drainage during the first 28 days of follow-up. If you censor patients at death, then you can estimate the hazard ratio for treatment on the cause-specific hazard of time to unassisted breather or pneumothorax requiring chest tube drainage. Can the authors further explore this and make their arguments more precise. A good reference for solutions to the competing risk problem is: Varadhan R, Weiss CO, Segal JB, Wu A, Scharfstein DO, Boyd C. Evaluating health outcomes in the presence of competing risks: A review of

	statistical methods and clinical applications. Medical Care. 2010; 48:S96-S105. (AHRQ Contract No. 290-05-0034) 4. *Potential effect modifiers and hypotheses:* The authors should make it clear in the second sentence of this paragraph that the potential effect modifiers are all baseline patient characteristics. 5. *Potential effect modifiers and hypotheses:* This reviewer prefers to see the hypotheses related to the potential effect modifiers stated in the introduction instead of in the methods section. 6. *Multivariable fractional polynomial interaction (MFPI) procedure:* Several key issues: a. I believe the utility of methodology papers like these is that they provide some good non-technical and intuition for complicated statistical methods. In this paragraph, I would love to see the authors provide a description of the multivariable fractional polynomial. Specifically, having chosen degree 2, the authors should provide a more complete description of what these functions are. In addition, in this paragraph the authors need to provide the statistical criteria (e.g. deviance, AIC or BIC) that was used to select the "best" multivariable fractional polynomial for each potential effect modifier. b. Please be more precise in defining the study specific likelihood ratio test for the interaction; these tests are not intuitive to most clinicians. c. It needs to be explained how you are estimating the TEF after adjusting for the potential confounders; i.e. are the TEFs presented in the figures based on setting the values of the potential confounders to 0? This is important as the TEFs can shift up or down depending on the value of the potential confounding variables. Please clarify. d. It should be stated in this paragraph that the confidence intervals represent pointwise confidence intervals. 7. *Averaging the functions of individual studies – meta-analysis:* a. The authors state "As averaging is done point-wise, distribution of events and therefore also of the patient population influences the weights" The authors should re-write this sentence so that it reads more clearly. The weights are calculated for each study incorporating the study specific estimate of the TEF with associated variance at each value of the potential effect modifier; the estimate and associated variance is a function of the number of events for patients with values of the potential effect modifier within a close range of the specific value. b. The authors should better justify the study-specific hypothesis test for the effect modification given that they are not providing the significance of the averaged function. In addition, it is not clear why the hypothesis test for the effect modification is not completed due to "treatment effects were non-linear". Can the authors justify this? This seems to be the entire goal of using the MFPI approach. c. Can the authors clarify whether or not they conducted these analyses separately for each potential effect modifier and outcome? 8. *Missing values and influential points:* Having the missing values for the potential effect modifiers is a complicating factor. I wonder if the results change when applying a simple mode/median

	replacement. In my opinion, this would be a simpler strategy in this case since the focus of the manuscript is on the MFPI / TEF approach. If the authors want to stick with the missing data imputation, a more complete description of the imputation approach should be included in an appendix; e.g. what models were specified? What if any additional variable were used in the imputation. 9. Results: Table 1: It would be helpful to see Q1/Q3 or min/max values for the potential effect modifiers. Perhaps including this in an appendix would be helpful for the reader to better understand the range for the potential effect modifiers. 10. Results: Second paragraph: The authors state “This may indicate that there are few, if any, differential interaction effects, but it may also be a result of low power to detect interactions in individual studies” Please modify this sentence as “interaction” defines differential effects. 11. Results: Third paragraph: The authors need to add a sentence to introduce how they plan to present the findings in Figure 2. Something to the effect: In what follows, we will describe the relationship between the treatment effect comparing high vs. low PEEP and each of the four potential effect modifiers. Then list some key items that are critical to understand figure 2 including an interpretation of the dashed vertical lines (this should also be included in the figure legend as well as for the on-line appendix figures 1 – 3). 12. On-line appendix figures: readers may be interested in the parameters selected for the MFPI models for each outcome/modifier combination. Since you are including these figures in the appendix, there should be room to include that information in the figure legend. 13. Results: this reviewer would prefer to see Figure 3 moved to an appendix to make more room in the results section to discuss the key features of Figure 2. 14. *Interaction with body mass index:* I disagree with the assessment provided in this paragraph. The authors suggest that higher PEEP may be beneficial for 60 day hospital mortality for patients with BMI > 35; however, the information in this region is very limited (as the authors go on to say). So why not just point out to the reader that the estimate of the averaged TEFs are negative for BMI > 35 but that these estimates should be judged with caution due to the wide CIs. In addition, the summary of the outcome: time to unassisted breathing is not accurate. In these graphs the main effect of BMI on time to unassisted breathing cannot be assessed; you are only quantifying the interaction. 15. *Interaction with PaO2/FiO2:* ”The averaged TEFs for PaO2/FiO2 suggest that patients with values below 150 but 100 mmHg above benefit” should be re-written to read “The averaged TEFs for PaO2/FiO2 suggest that patients with values below 150 but **above** 100 mmHg **may** benefit” AND “however, CIs at both ends are very wide and the functions for odds and hazard ratios hardly exclude ‘1’ for any value, leaving a high degree of uncertainty.” should be rewritten, e.g. “ however, the confidence intervals in this range

	are wide and barely exclude the odds and hazard ratio of 1 indicating no treatment effect." 16. *Summary of findings:* The conclusions here should be tampered by the fact that the statistical confidence is weak and include a discussion of the potential pool of study participants with values of the effect modifiers in these ranges; i.e. are these even common patients? And among these patients, the numbers of events are low based on the confidence limits. 17. *Comparison to the original analysis:* Given that you are describing the analysis and presenting the results of statistical tests for interactions, it would be beneficial to also include those p-value estimates for the meta-analysis with MFPI.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1
The authors present the results of a multivariable fractional polynomial interaction (MFPI) approach investigating the interactions between 4 continuous variables and the PEEP setting strategy, based on individual data from 3 major RCTs. They mainly found a non-linear effect modification of high PEEP by P/F ratio with reduced mortality for moderately severe ARDS patients; with a potent additional benefit of high PEEP in patients with high BMI. This is clearly a major topic from a clinical point of view. Obviously, the results extend our knowledge on this important topic, as compared to previous results published by the same group using pre-defined thresholds. However, as acknowledged by the authors, the results deserve further confirmations in external samples before modification of the current ventilator strategies in ARDS. Finally, the paper is well-written and I have very few comments.
Major comment

1. I have not sufficient knowledge of the statistical methodology for appreciating if the high percentage of missing values can affect the overall results, and if the imputation method is methodologically adequate. A statistical advice could be of value.

REPLY: Thank you for raising this important point. Missing values can indeed affect results by e.g. limiting the statistical power or even more important by introducing a selection bias. To address these potential caveats in our analyses, in our protocol we a priori specified to use multiple imputations by chained equation techniques to impute the missing data. As pre-specified, all analyses were based on the first imputed data set. The statistician who conducted the multiple imputations (PR) is very experienced in this field and has also written software packages for multiple imputations.
To provide the reader with more details about the multiple imputations, we now provide more information in the Appendix (please also see our reply to reviewer #2, comment 8):

"For the imputation of missing values we used multiple imputations by chained equations. This technique uses switching regression, which is an iterative multivariable regression technique and implemented in the STATA software with the ice command (Royston P. 2005. Multiple imputation of missing values: update. Stata Journal 5: 188-201; Royston P. 2005. Multiple imputation of missing values: update of ice. Stata Journal 5: 527-536.; Royston P. 2007. Multiple imputation of missing values: further update of ice, with an emphasis on interval censoring. Stata Journal 7: 445-464.). The following variables were considered in the chained equations to impute missing values: BMI (log-transformed), respiratory compliance, oxygenation index (log-transformed), adjusted PaO2/FiO2 (log-transformed), probability of death (inverse normal transformed), respiratory rate, plateau pressure at baseline, minute ventilation at baseline (log-transformed), tidal volume at baseline (log-transformed),

age, number of days on ICU before randomization, number of days of the ventilator, number of days until death in hospital, number of additional failed organs at baseline, set baseline PEEP, direct lung injury at baseline, presence of ARDS, gender, presence of severe sepsis, trial number, censoring code for death, censoring code for coming off the ventilator, group assignment, rescue therapy, death after rescue therapy, therapy with muscle relaxation, therapy with vasopressor support, therapy with steroids. According to the nature of the variable, linear regression, logistic regression or ordinal regression was used in the respective equations. We used the first imputed dataset for our analyses."

Minor comments
1.
Please check affiliation of Prof Laurent Brochard.

REPLY: We have checked and updated it.

2. Since the data are not really new, Figure 1 could perhaps be deleted.

REPLY: We agree that the data are not new, however, we think that Figure 1 may still be helpful for the reader, and we have a slight preference keeping it in the manuscript. However, it is also not critical to delete the figure and give the main information in the text. We are pleased to follow the advice of the Editor.

3. As driving pressure is known as the ventilation variable that best stratified the mortality risk, it could be interesting for the reader to know if it could be possible to include this important parameter in the MFPI approach.

REPLY: Thank you for this suggestion. Driving pressure is correlated with the investigated ventilation variables which makes it difficult/not useful to additionally adjust for it in our analyses. More important, driving pressure is not a patient baseline characteristic, but rather a feature of the intervention in combination with the baseline lung condition of individual patients. In subgroup/interaction analysis it is appropriate to exclusively stick to patient baseline characteristics as potential effect modifiers.

Reviewer #2
The authors present an interesting re-analysis of an individual patient data meta-analysis of three randomized trials assessing the benefits of high vs. low PEEP. In meta-analysis, potential effect modifiers can start to be explored by pooling information across trials; however, it is often common to dichotomize continuous candidate effect modifiers to provide simplicity of interpretation. The authors discuss a recent approach where the treatment effect is estimated as a smooth function of the candidate continuous effect modifier within each study and then these functions are pooled across studies using weight functions defined for each study across the range of the candidate continuous effect modifiers.
This paper would contribute the literature in two ways: a) this paper provides another example of how to implement and interpret the findings of a meta-analysis using fractional polynomials to estimate a treatment effect as a function of a continuous modifier, b) this paper contributes additional data beyond the original meta-analysis conducted on the three trials and offers additional insight into hypothesize effect modifiers for the effect of PEEP in patients with ARDS.

REPLY: Thanks, just a minor clarification which was not well expressed in our Introduction. The text is now slightly modified.
In this paper we provide results of the first ever planned example of such an extended analysis for continuous effect modifiers. Two of us (PR and WS) have developed the MFPI approach and an approach for a meta-analysis of continuous functions (refs 13 and 17), the two methods combined here. In 2012 four of us have registered the ICEM protocol which was extended to a publication (ref

19). Because of various reasons we needed some time to finalize the analysis and to write this paper. Meanwhile another group (ref 35) has partly followed our approach and has published a similar kind of analysis by using software provided by PR. We are not aware of any other analysis of this kind. Two of us (WS and PR) have started to work on a methodological paper for such analyses. In the protocol of the ICEM study we restricted to key issues of methods familiar to us, however we are well aware that further methodological research is required when assessing the potential effect of continuous treatment effect modifiers in a meta-analysis.

My points below are places in the manuscript where I feel the authors can strengthen the applicability of the manuscript for ARDS researchers who may not be statisticians and also places where extra information or a clarification of language would improve the content of the manuscript.

1. Introduction: Second paragraph: The authors should better motivate/describe the Sauerbrei and Royston approach for the meta-analysis. The relevance of the discussion in this paragraph is not tied very well in to the meta-analytic approach.

REPLY: We now describe the motivation for the MFPI approach in meta-analysis in more details as suggested in the Introduction section (page 6):

“Analyses of interactions are not only interesting for single trials, but also highly relevant for meta-analyses. Sutton et al have provided a comprehensive overview of many issues and developments in meta-analysis in the early years of this century [16]. However, approaches to combine results from several studies in a summary estimate of the functional relationship between a continuous covariate and an outcome have received more attention only in recent years. In the context of continuous risk or prognostic factors, Sauerbrei and Royston have proposed a new strategy for meta-analysis of functions, provided individual participants data (IPD) is available for the estimation of functions in each of the studies [17]. Obviously, this meta-analysis approach can also be used to derive an averaged treatment effect function summarizing TEFs from several MFPI analyses of randomized trials. To improve on several critical issues from cut-point based meta-analyses, we started the ICEM project. The key ingredients were all available: IPD from all randomized trials for a specific treatment comparison, the MFPI approach to estimate a TEF in each study, the meta-analysis approach to average several TEFs across studies and the required software [18]. In 2012 four of us have registered the ICEM protocol which was extended to a publication [19].”

The following two issues belong together – a reply is provided after point 3.

2. Clinical outcomes: First paragraph: The authors should be more precise in defining the statistical approach and models for the time to event analyses. Specifically, the author state “Patients who died before achieving unassisted breathing within the first 28 days were censored at the day of death. With this procedure we circumvented the problem of competing risk in intensive care trials [22] in the analysis of this outcome.” However, this procedure does not circumvent the competing risk issue, it modifies the interpretation of the standard time-to-event analysis to a cause-specific model. This should be more precisely defined for those applicable time-to-event outcomes.

3. Clinical outcomes: Second paragraph: I’m not entirely clear why the considerations for the competing risk of death differ for the time to unassisted breathing and pneumothorax requiring chest tube drainage during the first 28 days of follow-up. If you censor patients at death, then you can estimate the hazard ratio for treatment on the cause-specific hazard of time to unassisted breather or pneumothorax requiring chest tube drainage. Can the authors further explore this and make their arguments more precise. A good reference for solutions to the competing risk problem is: Varadhan R, Weiss CO, Segal JB, Wu A, Scharfstein DO, Boyd C. Evaluating health outcomes in the presence of competing risks: A review of statistical methods and clinical applications. Medical Care. 2010;

48:S96-S105. (AHRQ Contract No. 290-05-0034)

REPLY: We agree with the comment of the reviewer regarding the wording of competing risks of death in the Methods section which we have improved in the revised version of the manuscript. However, since we are interested in the direct (causal) effect of higher vs lower PEEP on the cause-specific hazard of unassisted breathing, the cause-specific Cox model is an appropriate choice for this question [Refs 25, 26]. The issue of model choice in the presence of competing risks depends on the context of the scientific question and is extensively described in the cited references. We have re-worded the methods section regarding this issue as follows:

“Because of differential follow-up across trials for this outcome beyond day 28 and the fact that the intervention effect is likely to precede day 28, as in the original meta-analysis we administratively censored patients at day 28 [6]. Patients who died before achieving unassisted breathing within the first 28 days were censored at the day of death. We are aware of the problem of competing risks in intensive care trials [24], but the MFPI approach has yet not been adapted for a competing risk analysis. To address this, we therefore used a cause-specific Cox model [25, 26] to assess the direct effect of higher versus lower PEEP on the cause-specific hazard of unassisted breathing.”

4. Potential effect modifiers and hypotheses: The authors should make it clear in the second sentence of this paragraph that the potential effect modifiers are all baseline patient characteristics.

REPLY: We have now re-worded this introducing sentence of this chapter clarifying that all potential effect modifiers are baseline patient characteristics:

“All of the following four potential effect modifiers were baseline patient characteristics measured at randomization and the analyses for potential interaction were all pre-specified in our published protocol.”

5. Potential effect modifiers and hypotheses: This reviewer prefers to see the hypotheses related to the potential effect modifiers stated in the introduction instead of in the methods section.

REPLY: We now provide the hypotheses in the last paragraph of the Introduction section as recommended:

“We hypothesized that patients with low PaO2/FiO2 or high oxygenation index and still good respiratory system compliance at baseline (moderate ARDS) have most recruitable lung units and would therefore benefit most from higher levels of PEEP. For patients at either end of the spectrum (mild ARDS or very severe ARDS) higher PEEP might not provide any benefit. We did not anticipate any specific direction of interaction effect for BMI.”

6. Multivariable fractional polynomial interaction (MFPI) procedure: Several key issues:

a. I believe the utility of methodology papers like these is that they provide some good non-technical and intuition for complicated statistical methods. In this paragraph, I would love to see the authors provide a description of the multivariable fractional polynomial. Specifically, having chosen degree 2, the authors should provide a more complete description of what these functions are. In addition, in this paragraph the authors need to provide the statistical criteria (e.g. deviance, AIC or BIC) that was used to select the “best” multivariable fractional polynomial for each potential effect modifier.

REPLY: We fully understand that you would like to see many more details about the methodology, but first we would like to clarify that this is a clinical paper trying to provide additional clinical information for a relevant issue of treating patients. To cope with the clinical questions, we combined two recent

methodological approaches. Unfortunately, our original Introduction was not sufficiently clear, but we hope having clarified this (see our reply to your first point).
All details concerning selection of treatment effect functions and averaging them have been published in methodological papers and cannot be repeated here. Having a methodological paper about combining MFPI and the meta-analysis approach would have been very helpful to better explain our analysis, but the started manuscript needs much more work and will not be available before the end of the year.
To select a specific FP function, you need to define a significance level (we used 5%) and the FP algorithm uses a closed test procedure to determine whether a variable has an effect and if the suitable function is linear, FP1 or FP2. MFPI is an extension and some additional issues are relevant. Writing the protocol for this study we had decided to use only FP2 functions for the analysis of interactions and select the two power terms which fit the data best. We have slightly extended the text and added the reference to a paper describing all MFPI details. In addition, we have added the reference to a new website about MFP modelling. The basic ideas are explained and references to many papers could help to better understand our approach.

b. Please be more precise in defining the study specific likelihood ratio test for the interaction; these tests are not intuitive to most clinicians.

REPLY: We agree that these tests are not intuitive to most clinicians but a suitable explanation is beyond the scope of this paper.

a. It needs to be explained how you are estimating the TEF after adjusting for the potential confounders; i.e. are the TEFs presented in the figures based on setting the values of the potential confounders to 0? This is important as the TEFs can shift up or down depending on the value of the potential confounding variables. Please clarify.

REPLY: We have added more details to the description of MFPI and we give all estimates of the derived main effect models in Table 1 to 4 in the appendix.

b. It should be stated in this paragraph that the confidence intervals represent pointwise confidence intervals.

REPLY: We have now added this information as recommended.

7. Averaging the functions of individual studies – meta-analysis:

a. The authors state “As averaging is done point-wise, distribution of events and therefore also of the patient population influences the weights” The authors should re-write this sentence so that it reads more clearly. The weights are calculated for each study incorporating the study specific estimate of the TEF with associated variance at each value of the potential effect modifier; the estimate and associated variance is a function of the number of events for patients with values of the potential effect modifier within a close range of the specific value.

REPLY: We agree that this part was a bit short and we have extended it. However, detailed explanations are beyond the scope of this paper. Interested readers can find more details in section 4.2. of ref 17.

b. The authors should better justify the study-specific hypothesis test for the effect modification given that they are not providing the significance of the averaged function. In addition, it is not clear why the hypothesis test for the effect modification is not completed due to “treatment effects were non-linear”. Can the authors justify this? This seems to be the entire goal of using the MFPI approach.

REPLY: P-values of study specific tests are given for information but they are not relevant for the approach and can be deleted. Happy to do it if you or the Editors prefers it. Obviously, RCTs are not powered to find interactions and p-values from study specific test are usually large. Dependent on selected power terms, treatment effect may simply be linear (the easiest case) but often non-linear functions will be selected. In such cases treatment effect functions are also non-linear.

c. Can the authors clarify whether or not they conducted these analyses separately for each potential effect modifier and outcome?

REPLY: We conducted the analysis for each potential effect modifier and outcome. A 'pair' of a potential modifier (e.g. BMI) and an outcome (e.g. in-hospital mortality) was considered as one investigation. In total, with four potential modifiers and three outcomes we had twelve investigations. We now clarify this in the Methods Section (Potential effect modifiers and investigations):

"A 'pair' of a potential modifier (e.g. BMI) and an outcome (e.g. in-hospital mortality) was considered as one investigation. In total, with four potential modifiers and three outcomes we had twelve investigations for each trial separately and for the meta-analysis."

8. Missing values and influential points: Having the missing values for the potential effect modifiers is a complicating factor. I wonder if the results change when applying a simple mode/median replacement. In my opinion, this would be a simpler strategy in this case since the focus of the manuscript is on the MFPI / TEF approach. If the authors want to stick with the missing data imputation, a more complete description of the imputation approach should be included in an appendix; e.g. what models were specified? What if any additional variable were used in the imputation.

REPLY: We agree that handling missing data is a crucial issue. We have pre-specified our approach of multiple imputations in our published protocol and we feel that it is the most appropriate approach. However, we agree that more details on the models could be helpful for the reader, therefore, also considering the comment by reviewer 1 (comment 1), we now provide more details about the technique and models used for the multiple imputations in the Appendix:

"For the imputation of missing values we used multiple imputations by chained equations. This technique uses switching regression, which is an iterative multivariable regression technique and implemented in the STATA software with the ice command (Royston P. 2005. Multiple imputation of missing values: update. Stata Journal 5: 188-201; Royston P. 2005. Multiple imputation of missing values: update of ice. Stata Journal 5: 527-536.; Royston P. 2007. Multiple imputation of missing values: further update of ice, with an emphasis on interval censoring. Stata Journal 7: 445-464.). The following 29 variables were considered in the chained equations to impute missing values: BMI (log-transformed), respiratory compliance, oxygenation index (log-transformed), adjusted PaO2/FiO2 (log-transformed), probability of death (inverse normal transformed), respiratory rate, plateau pressure at baseline, minute ventilation at baseline (log-transformed), tidal volume at baseline (log-transformed), age, number of days on ICU before randomization, number of days of the ventilator, number of days until death in hospital, number of additional failed organs at baseline, set baseline PEEP, direct lung injury at baseline, presence of ARDS, gender, presence of severe sepsis, trial number, censoring code for death, censoring code for coming off the ventilator, group assignment, rescue therapy, death after rescue therapy, therapy with muscle relaxation, therapy with vasopressor support, therapy with steroids. According to the nature of the variable, linear regression, logistic regression or ordinal regression was used in the respective equations. We used the first imputed dataset for our analyses."

9. Results: Table 1: It would be helpful to see Q1/Q3 or min/max values for the potential effect modifiers. Perhaps including this in an appendix would be helpful for the reader to better understand the range for the potential effect modifiers.

REPLY: We now provide the inter-quartile range for the potential effect modifiers in table 1.

10. Results: Second paragraph: The authors state “This may indicate that there are few, if any, differential interaction effects, but it may also be a result of low power to detect interactions in individual studies” Please modify this sentence as “interaction” defines differential effects.

REPLY: We have re-worded this sentence for clarification to:

“This may indicate that there are few, if any, interaction effects, but it may also be a result of low power to detect interactions in individual studies.”

11. Results: Third paragraph: The authors need to add a sentence to introduce how they plan to present the findings in Figure 2. Something to the effect: In what follows, we will describe the relationship between the treatment effect comparing high vs. low PEEP and each of the four potential effect modifiers. Then list some key items that are critical to understand figure 2 including an interpretation of the dashed vertical lines (this should also be included in the figure legend as well as for the on-line appendix figures 1 – 3).

REPLY: We agree that the way of presenting Figure 2 requires a little more detail. The key elements of the figures are already provided in the legends of the figures (including those in the appendix), therefore we suggest keeping the legends as they are. To provide the reader with a better description, we have now added a separate paragraph at the beginning of the results section (Interpretation of TEFs):

“The respective TEF graphs show the relationship between the continuous patient characteristic (on the X-axis, e.g. BMI) and the benefit/harm (e.g. expressed as OR on the Y-axis) from the intervention (higher versus lower PEEP). If the average effect (dashed bold line) is below the horizontal line, this suggests a treatment benefit from higher PEEP and a detrimental effect if it is above. It is the opposite for the endpoint time-to-unassisted breathing, because here, the outcome of interest is the “positive” event coming off the ventilator. Therefore, an OR of e.g. 1.3 expresses the increased chance to come off the ventilator with higher PEEP. The thin dashed lines represent the point-wise upper and lower 95% CI limit. Therefore, at those parts where the 95% CI includes the horizontal line, it is uncertain whether the suggested benefit or harm is real. The dashed vertical lines indicate the 5% and 95% centile of the data of the continuous predictors.”

12. On-line appendix figures: readers may be interested in the parameters selected for the MFPI models for each outcome/modifier combination. Since you are including these figures in the appendix, there should be room to include that information in the figure legend.

REPLY: We now provide more information about the parameters in the Appendix Tables 1 to 4.

13. Results: this reviewer would prefer to see Figure 3 moved to an appendix to make more room in the results section to discuss the key features of Figure 2.

REPLY: We agree that Figure 2 displays the main objective of our analyses. However, averaging of functions from several MFPI analyses is a new approach in meta-analysis and the contribution of each study to the averaged TEF is likely not familiar to many readers, therefore we suggest leaving Figure 3 in the main manuscript. Of course, we don’t insist on this and would be happy to follow the

editor's suggestion.

14. Interaction with body mass index: I disagree with the assessment provided in this paragraph. The authors suggest that higher PEEP may be beneficial for 60 day hospital mortality for patients with BMI > 35; however, the information in this region is very limited (as the authors go on to say). So why not just point out to the reader that the estimate of the averaged TEFs are negative for BMI > 35 but that these estimates should be judged with caution due to the wide CIs. In addition, the summary of the outcome: time to unassisted breathing is not accurate. In these graphs the main effect of BMI on time to unassisted breathing cannot be assessed; you are only quantifying the interaction.

REPLY: We agree that there is considerable increase in uncertainty once the value of BMI 35 is passed. To avoid any spin from this analysis, we have tuned down the wording as recommended:

"Regarding all three outcomes, the TEFs do not suggest particular interactions. However, especially regarding hospital mortality and time-to-death, the 95% CIs are considerably wide at both ends, which leaves much uncertainty about the interaction effect."

15. Interaction with PaO2/FiO2: "The averaged TEFs for PaO2/FiO2 suggest that patients with values below 150 but 100 mmHg above benefit" should be re-written to read "The averaged TEFs for PaO2/FiO2 suggest that patients with values below 150 but above 100 mmHg may benefit" AND "however, CIs at both ends are very wide and the functions for odds and hazard ratios hardly exclude '1' for any value, leaving a high degree of uncertainty." should be rewritten, e.g. " however, the confidence intervals in this range are wide and barely exclude the odds and hazard ratio of 1 indicating no treatment effect."

REPLY: We agree that this section requires more clarity. However, we do not entirely agree that the proposed statement "indicating no treatment effect" reflects the quantitative assessment of the TEF curve. We have re-worded:

"They suggest that patients with values below 150 but above 100 mmHg (moderate ARDS) may benefit with respect to all three outcomes, however, the confidence intervals in this range are still wide and barely exclude the odds and hazard ratio of 1 reflecting some remaining uncertainty about the treatment modifying effect of PaO2/FiO2. At both ends, 95% CIs are very wide and the functions for odds and hazard ratios hardly exclude '1' for any value, leaving a high degree of uncertainty."

16. Summary of findings: The conclusions here should be tampered by the fact that the statistical confidence is weak and include a discussion of the potential pool of study participants with values of the effect modifiers in these ranges; i.e. are these even common patients? And among these patients, the numbers of events are low based on the confidence limits.

REPLY: We have re-worded the summary of findings considering the low confidence in the treatment modifying effects and also commented on the pool of study participants:

"In this meta-analysis of three randomized trials, we have used a novel approach (MFPI) to investigate interactions between continuous baseline patient characteristics with two ventilation strategies (high versus lower PEEP). Although the statistical confidence in most interaction effects is very weak, the visual inspection of averaged TEFs suggests that patients with moderate ARDS may benefit from higher PEEP ventilation with respect to hospital mortality, time-to-death, and time-to-unassisted breathing. Patients with severe obesity were excluded from all included trials, therefore although the TEF for BMI may suggest mortality reduction with higher PEEP ventilation in patients with very high BMI, no conclusions for the clinical practise can be made, because the uncertainty about this potential interaction is very large. Caused by inclusion and exclusion criteria of the

individual studies, estimates in the outside range have the flavour of an extrapolation into regions where real data is missing."

17. Comparison to the original analysis: Given that you are describing the analysis and presenting the results of statistical tests for interactions, it would be beneficial to also include those p-value estimates for the meta-analysis with MFPI.

REPLY: We now provide the P-value (interaction analysis) from the original meta-analysis (not using MFPI) in the discussion section. However, we did not calculate p-values for the averaged TEFs. Several possibilities to calculate p-values exist and further methodological work is required before we can recommend a suitable p-value. In addition, we feel that the visual assessment of the TEF provides much more information and that observers can infer 'statistical significance' from the positions of the 95% CIs. However, we are also aware that the CIs are too small, specifically in the outside range of the curves. Some more information can be found on the website (Ref 27).

VERSION 2 – REVIEW

REVIEWER	Jean-Luc Diehl Critical care Medicine Depatment, Georges pompidou European Hospital INSERM UMR_S1140 Paris-Descartes University Paris. France.
REVIEW RETURNED	18-Jun-2016

GENERAL COMMENTS	The authors adequately answered my previous questions/comments. I consider that the manuscript is now suitable for publication. Thanks for giving me the opportunity to review this important study.

REVIEWER	Elizabeth Colantuoni Bloomberg School of Public Health, Johns Hopkins University, USA
REVIEW RETURNED	30-Jun-2016

GENERAL COMMENTS	I thank the authors for so carefully and thoroughly responding to my many questions/comments. The revisions have strengthened the manuscript in my opinion. My one point that I don't think was adequately addressed concerns the outcome of pneumothorax. The authors have not entirely described why this outcome is different in regard to the competing risk of mortality compared to time to unassisted breathing. I would recommend removing mention of this outcome from the methods section; which will not substantially effect the manuscript and will limit any confusion regarding this outcome. I will leave the editor to decide over the inclusion or exclusion of Figure 3. I think the methods description has been modified in a way to make this more intuitive for the reader.