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)	Determining a safe upper limit of oxygen supplementation for adult patients: a systematic review
AUTHORS	Lassen, Mathilde Languille; Risgaard, Bjarke; Baekgaard, Josefine; Rasmussen, Lars

VERSION 1 – REVIEW

REVIEWER	Ore, Timothy Commission for Hospital Improvement, Department of Health
REVIEW RETURNED	11-Oct-2020

GENERAL COMMENTS	With only 12 studies reviewed, and given the variability of the results collectively, the limitations section of the manuscript needs further strengthening, particularly to give assurance that inclusion of "missed' studies is unlikely to substantially change final conclusions drawn. To avoid confusion, I suggest the sentence in the conclusion section of the Abstract and at the end of Discussion (page 13), beginning "In this systematic review we found that there was inadequate evidence to identify a safer upper dosage of oxygen....." be the same. Have a heading for Statistical Analysis in the Methods section that details the stats used, including choice of significance levels.

REVIEWER	Godet, Thomas Centre Hospitalier Universitaire Clermont-Ferrand, Departement de Médecine PériOpératoire
REVIEW RETURNED	01-Nov-2020

GENERAL COMMENTS	Dear Authors, I would like to thank you for this meta-analysis dealing with a subject of high interest: which level of FiO2 should be used in critical care and preoperative medicine, in terms of complications (atelectasis, duration of mechanical ventilation, pneumonia and ARDS). Despite an excellent writing and ease of reading, very few information are given by your study because of included studies heterogeneity and conclusions are to my point of view hazardous. Your study does not offer physicians answers to main question. I have several major concerns, with several points needing to be specified: 1. Selected studies have been conducted in different settings (elective surgery or critical care). Comparison of data are consequently hazardous to my point of view;

	2. Different outcomes' evaluation delays create confounding analyses and conclusions hazardous; 3. Atelectasis definitions are different between studies. Some used CT scan measurements, others standard X-rays... Finally, pooling of those studies renders conclusions at high risk of bias; 4. ARDS must be defined according to Berlin definition. Moreover, the absence of data on ARDS timing (early or late onset) could have had a large impact on outcome analysis; 5. Definitions of pneumonia must be given in the manuscript (ATS definition?) in relation with those used in selected studies, since a related high risk of bias rises; 6. A large overlap in FiO2 levels between the two groups is a major confounder. I wonder how was the level of 0.8 determined to conclude this is the upper acceptable level of FIO2? I also have few minor comments : 1. Page 9: Numbers must have identical significant figures; 2. Page 10, Line 6 : I would suggest to change “bigger” to “larger” Finally, it would be interesting to present a funnel plot for publication bias and I2 (heterogeneity) must be detailed in main text.

REVIEWER	Knapp, Guido Technicsche Universitat Dortmund
REVIEW RETURNED	16-Feb-2021

GENERAL COMMENTS	The authors did a systematic review for determining a safe upper limit of oxygen supplementation, which was expeceted to be a descriptive summary. In their analysis, they followed the PRISMA guideline. In total, the authors considered four diffeerent endpoints. For two of them, a meta-analysis was possible. The authors did not clearly described the meta-analysis model used. Based on the section on study charecteristics, the random-effects meta-analysis should be used. For atelectasis, only four out of eight studies can beu used and the authors got a significant result in favour of the low group. However, according to Figure 2, the authors used the Mantel-Haenszel method for combining the relative risks, which is a method for the ficxd-effect meta-analysis model. The authors should redo the meta-analysis in a random-effects model. to confirm the significant result prominently mentioned in the abstract. For pneumonia, although four studies with relative risks are reported but no meta-analysis carried out. Is there any reason?

VERSION 1 – AUTHOR RESPONSE

#Reviewer 1: Dr. Timothy Ore, Commission for Hospital Improvement

With only 12 studies reviewed, and given the variability of the results collectively, the limitations

section of the manuscript needs further strengthening, particularly to give assurance that inclusion of "missed' studies is unlikely to substantially change final conclusions drawn.

Thank you for this comment. Unfortunately there is always a risk of missed studies. However, we made a systematic search with a predefined search string, and we also screened reference lists of included studies. Due to the heterogeneity of the 12 studies reviewed, we do not believe that any potentially missed studies could have changed our conclusion substantially. Even though the potential of missed studies is already mentioned in our limitations section, we have replaced it with the the following for further strengthening:

"Despite the systematic search with predefined search string, and screening of reference lists of included studies, there is always a possibility that our search did not identify all relevant studies. However, the heterogeneity of the 12 studies reviewed makes us believe that potentially missed studies would not change the conclusion substantially."

To avoid confusion, I suggest the sentence in the conclusion section of the Abstract and at the end of Discussion (page 13), beginning "In this systematic review we found that there was inadequate evidence to identify a safer upper dosage of oxygen....." be the same.

Thank you for pointing this out. This has been corrected.
"In this systematic review we found inadequate evidence to identify a safe upper dosage of oxygen"

Have a heading for Statistical Analysis in the Methods section that details the stats used, including choice of significance levels.

Thank you for this comment. We have followed the PRISMA-guidelines and used the suggested headings (http://www.prisma-statement.org/). If the Editors agree that there should be a heading called statistical analysis we are of course willing to do so. However, we should of course add the choice of significance levels as suggested, and therefore we have added the following to the section called "Summary measures and synthesis of results".
"Relative risks with 95% confidence intervals, was calculated in studies where this information was missing and the calculation was possible. The forest plot was made with a random-effects model.".

Regarding the choice of significance levels, we used the conventional cut-off point of 0.05, representing a 95% confidence interval.

#Reviewer 2: Dr. Thomas Godet, Centre Hospitalier Universitaire Clermont-Ferrand

1. Selected studies have been conducted in different settings (elective surgery or critical care). Comparison of data are consequently hazardous to my point of view;

Thank you for this comment. You are absolutely right. To accommodate this and other of the below mentioned concerns we have added the following to the Strengths and limitations section:
"The trials varied in patient groups, associated clinical care and disease severity. Furthermore, in some studies it is unclear when exactly the outcome of interest was measured (early or late onset of ARDS and timing of CT/X-ray for measuring the presence of atelectasis). It is also unclear how pneumonia was defined in the four studies reporting this outcome. Therefore conclusions should be drawn with caution."

2. Different outcomes' evaluation delays create confounding analyses and conclusions hazardous;

Thank you for pointing this out. We hope that the above-mentioned addition to the manuscript will accommodate your concerns regarding this subject.
We also looked into the studies reporting atelectasis and found the following:

Study	***Outcome evaluation***
Akca et al	*First postoperative day*
Asfar et al	*First 3 days*
Benoit et al	*When the patient was fully awake*
Ishii et al	*At arrival and 24 h after submission (oxygen supplementation given in between to scans)*
Lång et al	*Not documented*
Rothen et al	*During anaesthesia*
Staer et al	*Postoperatively, when relevant*
Suzuki et al	*Not documented*

To be thorough we have searched for studies reporting the period of time that atelectasis persist. According to a study by Eichenberger et al[1], atelectasis resolves within 24 h after laparoscopy in nonobese subjects. Another study by Lindberg et al[2] found that atelectases persist for 2 days after major surgery. Most of the included studies therefore evaluated the outcome within the timeline of atelectasis formation and persistence.

3. Atelectasis definitions are different between studies. Some used CT scan measurements, others standard X-rays... Finally, pooling of those studies renders conclusions at high risk of bias:

*Thank you for pointing this out. We agree on this comment and we have already pointed this out in the Discussion: "*Atelectasis was defined in different ways complicating the pooling of data and the possibility to undertake a meta-analysis"*.*

4. ARDS must be defined according to Berlin definition. Moreover, the absence of data on ARDS timing (early or late onset) could have had a large impact on outcome analysis;

Thank you for this comment. We agree that the most ideal way to define ARDS is according to the Berlin definition, however this was not an inclusion criterion (the inclusion criterion was "ARDS as defined by authors" – cf. our predefined protocol). In the study by Panwar et al they defined ARDS according to the Berlin definition, however in Lång et al this data was unavailable (as we have pointed out in the Discussion).
*In the section called "**Data collections and data items"**, the following was mentioned: "…as well as data on the predefined outcomes (atelectasis, pneumonia, ARDS) as defined by the authors"*
To accommodate your concerns and clarify the inclusion criteria we have added the following to the Eligibilty criteria section:
"Outcome: atelectasis, pneumonia, ARDS and duration of mechanical ventilation (as defined by authors)"

5. Definitions of pneumonia must be given in the manuscript (ATS definition?) in relation with those used in selected studies, since a related high risk of bias rises;

Thank you for this comment. We included studies reporting pneumonia as defined by authors (which we have now added more clearly to the section Eligibility criteria). Unfortunately the definition of pneumonia was unavailable from the selected studies.
In a future study, when more data is available, this would be a great idea to look further into.

6. A large overlap in FiO2 levels between the two groups is a major confounder. I wonder how was the level of 0.8 determined to conclude this is the upper acceptable level of FIO2?

Thank you for this comment. We understand why this needs clarification. The level of 0.8 was determined based on the studies in the forest plot, where the findings in three (Akca et al, Asfar et al and Staehr et al) suggested a benefit of an oxygen fraction below 0.8 while the results in the study by Lång et al were in favor of an oxygen fraction above 0.7.

I also have few minor comments :
1. Page 9: Numbers must have identical significant figures;

Thank you for this comment. The manuscript has been revised accordingly.

2. Page 10, Line 6 : I would suggest to change "bigger" to "larger"

Thank you for this comment. The manuscript has been revised accordingly.

Finally, it would be interesting to present a funnel plot for publication bias and I2 (heterogeneity) must be detailed in main text.

Thank you for this comment. We agree that I2 should be specified in the main text. We have therefore added the following to Results section:

"The heterogeneity (I2) of the meta-analysis presented in figure 2 is 31%, which corresponds to a moderate heterogeneity (Cochrane Handbook for Systematic Reviews of Intervention, section 9.5.2 Identifying and measuring heterogeneity)"

#Reviewer 3: Dr. Guido Knapp, Technicsche Universitat Dortmund.

In total, the authors considered four different endpoints. For two of them, a meta-analysis was possible. The authors did not clearly described the meta-analysis model used. Based on the section on study characteristics, the random-effects meta-analysis should be used. For atelectasis, only four out of eight studies can be used and the authors got a significant result in favour of the low group. However, according to Figure 2, the authors used the Mantel-Haenszel method for combining the relative risks, which is a method for the fixed-effect meta-analysis model. The authors should redo the meta-analysis in a random-effects model. to confirm the significant result prominently mentioned in the abstract.

Thank you for this comment. We agree on this excellent observation. We have now used a random-effect model and the Results section have been changed accordingly, leading to a Risk Ratio of 1.37 [0.95, 1.96].

For pneumonia, although four studies with relative risks are reported but no meta-analysis carried out. Is there any reason?

Thank you for this excellent suggestion. We have now added a forest plot using the data on pneumonia. To comment this new figure, we have added the following to the Results section:

"These studies are illustrated in the forest plot (figure 3), which shows a non-significant tendency that treatment with high FiO_2 was associated with higher risk of pneumonia, RR: 1.32 [0.65, 2.70] "

VERSION 2 – REVIEW

REVIEWER	Ore, Timothy Commission for Hospital Improvement, Department of Health
REVIEW RETURNED	07-Apr-2021

GENERAL COMMENTS	A good manuscript that could be improved by (a) indicating reasons why only two databases were used for the review and (b) the calculation of fail-safe number, to indicate the estimated quantum of studies that will be required to invalidate reported findings/inferences drawn. Regarding the former, in addition to MEDLINE and EMBASE, there are other great databases for the theme of the investigation; for example, EBSCO CINAHL Compete, Cochrane Library, Web of Science, LILACS, PubMed (non-MEDLINE records only) and UpToDate. Ignoring these sources need to be included in the discussion of the review limitations. Need to pay attention to spelling errors - hyperoxia/hypoxia, for instance.

REVIEWER	Knapp, Guido Technicsche Universitat Dortmund
REVIEW RETURNED	08-Apr-2021

GENERAL COMMENTS	The authors did the meta-analysis according to my prefious commnets. Only the heading "M.-H. Random" in the forest plots is still a bit confusing for me.

VERSION 2 – AUTHOR RESPONSE

#Reviewer 1: Dr. Timothy Ore, Commission for Hospital Improvement

Comments to the Author: A good manuscript that could be improved by

(a) indicating reasons why only two databases were used for the review

Thank you for this comment. We have now added this as a limitation, and also why this approach was chosen:

It is possible that more studies could have been found by searching in a wider set of databases. However, we chose the most commonly used databases MEDLINE and EMBASE, where the quality is known to be best and where most studies are found.

(b) the calculation of fail-safe number, to indicate the estimated quantum of studies that will be required to invalidate reported findings/inferences drawn.

Thank you for pointing this out. The calculation of fail-safe number is not used very often and not recommended by the Cochrane Collaboration (https://handbook-5-1.cochrane.org/chapter_10/10_4_4_3_fail_safe_n.htm) that specifically states the following: "the estimate

of fail-safe N is highly dependent on the mean intervention effect that is assumed for the unpublished studies (Iyengar 1988), and available methods lead to widely varying estimates of the number of additional studies (Becker 2005). The method also runs against the principle that in medical research in general, and systematic reviews in particular, one should concentrate on the size of the estimated intervention effect and the associated confidence intervals, rather than on whether the P value reaches a particular, arbitrary threshold, although related methods for effect sizes have also been proposed (Orwin 1983). Therefore this and related methods are not recommended for use in Cochrane reviews."

To accommodate your point of view, we would like to focus on the effect size. Therefore we added the following to the discussion:
Looking at figure 2, there is a relative risk of 1.37, which suggests a clinically relevant difference with less atelectasis with a lower oxygen fraction. However, the confidence interval is wide (0.95-1.96), indicating that more information is needed before any firm conclusions can be made.

Regarding the former, in addition to MEDLINE and EMBASE, there are other great databases for the theme of the investigation; for example, EBSCO CINAHL Compete, Cochrane Library, Web of Science, LILACS, PubMed (non-MEDLINE records only) and UpToDate. Ignoring these sources need to be included in the discussion of the review limitations.

Thank you for pointing this out. As mentioned above, we have now added this to the limitations.

Need to pay attention to spelling errors - hyperoxia/hypoxia, for instance.
Thank you for this comment. The manuscript has been revised accordingly.

#Reviewer 3: Dr. Guido Knapp, Technicsche Universitat Dortmund

Thank you for acknowledging this.
M.-H. Random is the model used for the meta-analyses. It is a random-effects model extension to the standard Mantel-Haenszel that is used by the program Review Manager. However, we understand the confusion, which is why we added the following to the figure caption:
Abbreviations: M.H. Random: Maentel-Haentzel Random effects model.

VERSION 3 – REVIEW

REVIEWER	Ore, Timothy Commission for Hospital Improvement, Department of Healths
REVIEW RETURNED	09-Jul-2021

GENERAL COMMENTS	Publication recommended.