PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

Title (Provisional)

Effect of subanesthetic dose of esketamine induction on quality of recovery from general anaesthesia in abdominal surgery: a propensity-score-matched retrospective study

Authors

Wang, Dongdong; Weng, Mengcao; Chen, Kunwei; Wu, Xiaojun; Xiao, Yuanfang; Wu, Yijie; Qian, Minyue; Lu, Zhongteng; Fang, Xiangming; Jin, Yue

---

VERSION 1 - REVIEW

Reviewer **1**

Name **ShangGuan, Wangning**

Affiliation **The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Department of Anesthesiology and Perioperative Medicine**

Date **04-Feb-2025**

COI **None**

---

1. Line 39, 20 VS 23, please add the unit of minutes.

2. Line 104, the exclusion criteria include "intraoperative hemodynamic fluctuations". In fact, most surgical procedures will have hemodynamic fluctuations inevitably.

3. Lines 175-176, regarding the sample size calculation, it is recommended to give specific "previous results" values to facilitate the verification of sample size.

4. Line 249, according to the Results section, subanesthetic doses of esketamine mainly reduce the recovery room stay time by affecting the postoperative extubation time, and it seems that it cannot be extended to affect the entire postoperative recovery time.

Reviewer	**2**
Name	**Dutta, Amitabh**
Affiliation	**Sir Ganga Ram Hospital, Anaesthesiology, Pain, and Perioperative Medicine**
Date	**17-Feb-2025**
COI	**None**

This Author has the following issues with the study context and philosophic take on the result/outcomes:

1. If sub-aesthetic doses of ketamine are to be studied then why chose s-ketamine? subanesthetic doses of standard ketamine could have been suffice as the side effects of ketamine become active at standard dose

2. The difference in recovery times, 2-3 minutes, although significant, doest not add much to the recovery room turnover, postoperative patient ambulation, and overall recovery from anesthesia. Therefore, the study results becomes redundant when seen on the ground of anesthesia practice.

Reviewer	**3**
Name	**Dack, Kyle**
Affiliation	**University of Bristol, MRC Integrative Epidemiology Unit**
Date	**26-Mar-2025**
COI	**None**

Statistical review summary:

Generally the methods and results are reported accurately. My only major concern is the propensity score matching – there are important details missing from the methods, as described below. Additionally, it is not evaluated, either through sensitivity analyses or just in the discussion – what is the risk of unmeasured confounding? The authors should judge this in the discussion because it is the key determinant of the validity of the results.

Abstract

1. Lines 25-28. These two sentences are repetitive and I suggest revising, perhaps remove the second sentence?

2. Line 29-30. Where were these patients recruited?

3. Methods sentence 3 and 5 are repetitive and I suggest merging.

4. Line 39. The units are missing – minutes?

5. I am a little confused about the naming of the different outcomes. The primary outcome is stated as PACU recovery time – this is not clearly defined, compared to the main methods which explains it is time to extubation. But in the abstract results, line 39 reports differences in "recovery time", while line 41 reports differences in "total PACU time" – what is the difference? There is also PACU observation time, which is not mentioned as a secondary outcome in the abstract methods.

I can see from the main methods section that there are a lot of secondary outcomes and maybe it is not possible to list all in the abstract. But the authors should at least ensure that any outcomes reported in the abstract results are first introduced in the abstract methods.

I also suggest the authors select a single phrasing for each outcome and check the manuscript to ensure it is named consistently throughout, and ensure all outcomes are named in the relevant positions.

Introduction

6. Clear and meets STROBE checklist criteria – no suggestions.

Methods

7. Line 96. Undergoing should be "underwent", or even better, "had".

8. Line 109. This sentence has unclear grammar and needs revising.

9. Line 153. The primary outcome is anesthesia recovery time (labelled T1), which is the cessation of anesthesia to extubation. The units of the outcome should be stated.

10. Line 175-176. I am not sure a sample size calculation is needed given that this is a retrospective analysis, and the authors cannot control the number of participants available. However, if the authors wish to report this, please add the expected effect size, variance, and the references from which this was taken.

11. Propensity score matching: I suggest moving this to separate subheading prior to the statistical analysis, because it is essentially a preparatory activity. There are also some details missing;

-The rationale for using propensity score matching rather than standard regression (perhaps could be added to the introduction)

-How were the matching variables chosen?

-Certain matching variables are ambiguous, e.g. surgical category (what categories), surgery duration (categorical or continuous?), medical history, please add details so these variables are better understood.

-It is stated that linear regression was used to estimate the matching, but typically PSM would use logistic regression to match receiving treatment 1 vs receiving treatment 2. R MatchIt also defaults to logistic regression. Could this be explained more – including all parameters specified in the MatchIt functions?

-How was the matching evaluated, and were any diagnostics/sensitivity analyses performed?

Results

12. Line 207-208. This approach to evaluating the PSM should be explained in the methods.

13. Figure 2. How was SMD calculated for categorical variables? This should be explained in the methods.

14. Figure 3. Distance should be defined somewhere.

15. Table 1. The categories for the matching variables should be explained in the methods. The descriptive statistics are not explained – suggest adding to the table label. E.g. 58.00 (45.00, 66.00) – median and IQR? I don't think z-statistics or chi-square statistics are necessary to report because the reader cannot interpret these – P and SMD should be enough. Removing those statistics might solve some of the line-wrapping issues.

16. Line 215. The p-value threshold will be clearer if added to line 212 after "differences"

17. Table 2. Data are presented as median/IQR or n/% - but which rows? I don't think "statistic" is needed? I think for full clarity, T2 and T3 should also be reported in the main text, to avoid reporting only significant results.

18. Table 3. Generally for all the tables I recommend reducing the line spacing to improve readability, and checking the alignment of values down columns.

19. The primary assumption of PSM is that there is no unmeasured confounding. Have the authors considered performing sensitivity analyses such as estimating E-

values to assess this? The discussion also should consider this qualitatively, currently the discussion does not mention confounding at all.

Discussion

20. I am unsure why the study being single center is a limitation – the reason should be explained, plus the impact on the results.

VERSION 1 - AUTHOR RESPONSE

Reviewer: 1

1. Line 39, 20 VS 23, please add the unit of minutes.

Response: Thank you for your suggestion. We have added the unit of minutes. (Page 2, Line 41)

2. Line 104, the exclusion criteria include "intraoperative hemodynamic fluctuations". In fact, most surgical procedures will have hemodynamic fluctuations inevitably.

Response: Thank you for your advice. The term 'intracellular haemodynamic fluctuations' refers to severe haemodynamic fluctuations with markedly unstable vital signs caused by massive bleeding. We have revised the manuscript. (Page 6, Line 109 – Page 7, Line 110)

3. Lines 175-176, regarding the sample size calculation, it is recommended to give specific "previous results" values to facilitate the verification of sample size.

Response: Thank you for your comments. We have revised the manuscript. (Page 11, Line 200-205)

Based on our previous results, the mean extubation time for both groups were 19.00±11.52 min and 22.15±14.42 min, respectively. A two-tailed test with α set at 0.05, 90% power and a sample size of 1:1 indicated that a minimum sample size of 361 participants per group was required. As PSM will be used for case selection, we included a larger sample size to ensure that the final number after PSM met the required

threshold.

4. Line 249, according to the Results section, subanesthetic doses of esketamine mainly reduce the recovery room stay time by affecting the postoperative extubation time, and it seems that it cannot be extended to affect the entire postoperative recovery time.

Response: Thank you for your advice. Anaesthesia recovery time (T1) is defined as the time from discontinuation of anaesthesia to extubation. We have changed the term to 'extubation time (T1)' for clarity. The use of subanaesthetic doses of esketamine for induction of anaesthesia significantly reduced the etubation time (T1). There was no statistical difference in PACU observation time (T2), probably due to the requirement for a minimun of 30 minutes post-extubation observation in the PACU. However, total PACU time was shorter in the esketamine group compared to the control group (62 minutes vs. 66 minutes, $p = 0.015$).

Reviewer: 2

1. If sub-aesthetic doses of ketamine are to be studied then why chose s-ketamine? subanesthetic doses of standard ketamine could have been suffice as the side effects of ketamine become active at standard dose

Response: Thank you for your suggestion. Ketamine is a commonly used anaesthetic in clinical practice, characterised by potent sedative and analgesic effects. It has been the subject of extensive research. However, its clinical use is limited by a number of adverse effects, including hallucinations, dizziness, delirium, nightmares and drowsiness.

Ketamine is a racemic mixture containing two optical isomers, S(+)-ketamine (esketamine) and R(-)-ketamine. The anaesthetic effect of esketamine is twice that of a racemic mixture and its potency is approximately three times that of (R)-ketamine [1]. In addition, esketamine has a lower incidence of adverse effects compared to racemate ketamine, which is a potential clinical advantage [2]. Given these advantages, the clinical use of esketamine has expanded. However, current research on esketamine remains relatively scarce compared to ketamine. Therefore, further studies on its clinical use are of considerable importance and may provide valuable insights for evidence-based clinical practice.

2. The difference in recovery times, 2-3 minutes, although significant, doest not add much to the recovery room turnover, postoperative patient ambulation, and overall recovery from anesthesia. Therefore, the study results become redundant when seen on the ground of anesthesia practice.

Response: Thank you for your comments. Although a 2-3 minutes reduction in recovery time may have limited clinical significance for individual patients, it can significantly reduce the use of PACU resources when applied to a high volume of surgeries.

Esketamine offers several advantages, including multi-receptor activity, combined sedative and analgesic effects, shorter recovery time and a lower incidence of adverse effects. Current perioperative pain management is shifting toward multimodal analgesia

with reduced opioid use. However, the impact of the use of multiple anaesthetic agents on postoperative recovery time and pain control remains controversial.

In this retrospective controlled trial, our results show that esketamine not only accelerates postoperative recovery, but also effectively reduces postoperative pain and opioid consumption. It also reduces the incidence of post-extubation respiratory depression without increasing adverse events such as PACU delirium, agitation, nausea and vomiting. These findings support the clinical value of esketamine. We therefore believe this study is of significant clinical importance.

Reference

1. Zeilhofer HU, Swandulla D, Geisslinger G, Brune K. Differential effects of ketamine enantiomers on NMDA receptor currents in cultured neurons. *Eur J Pharmacol.* 1992;213(1):155-158.

2. Wang J, Huang J, Yang S, et al. Pharmacokinetics and Safety of Esketamine in Chinese Patients Undergoing Painless Gastroscopy in Comparison with Ketamine: A Randomized, Open-Label Clinical Study. *Drug Des Devel Ther.* 2019;13:4135-4144.

Reviewer: 3

Statistical review summary:

Generally the methods and results are reported accurately. My only major concern is the propensity score matching – there are important details missing from the methods, as described below. Additionally, it is not evaluated, either through sensitivity analyses or just in the discussion – what is the risk of unmeasured confounding? The authors should judge this in the discussion because it is the key determinant of the validity of the results.

Response: Thank you for your comments. We have added more details in the manuscript, the rationale for using propensity score matching, how the matching variables were selected, the results of the multiple linear regression analysis, how the matching was assessed, etc. The detailed response can be found in questions 11, 12, and 19.

Abstract

1. Lines 25-28. These two sentences are repetitive and I suggest revising, perhaps remove the second sentence?

Response: Thank you for your suggestion. We have removed the second sentence.

2. Line 29-30. Where were these patients recruited?

Response: Thank you for your comments. Patients undergoing abdominal surgery with tracheal intubation in a tertiary academic hospital were retrospectively reviewed. We have added it to the manuscript. (Page 2, Line 30)

3. Methods sentence 3 and 5 are repetitive and I suggest merging.

Response: Thank you for your advice. We have merged these two sentences. (Page 2, Line 30)

4. Line 39. The units are missing – minutes?

Response: Thank you for your suggestion. We have added the unit of minutes. (Page 3,

Line 35-38)

5. I am a little confused about the naming of the different outcomes. The primary outcome is stated as PACU recovery time – this is not clearly defined, compared to the main methods which explains it is time to extubation. But in the abstract results, line 39 reports differences in "recovery time", while line 41 reports differences in "total PACU time" – what is the difference? There is also PACU observation time, which is not mentioned as a secondary outcome in the abstract methods.

Response: Your suggestions will be greatly appreciated.

The anaesthetic recovery time (T1) is originally defined as the time from discontinuation of anaesthetic to extubation; for clarity, this was revised to "extubation time (T1)". As all patients required observation after tracheal extubation, the time from extubation to PACU discharge was defined as the PACU observation time (T2). If patients experienced severe pain, vomiting, haemodynamic instability, respiratory depression or agitation after extubation, the PACU observation time was extended. Total PACU time (T3) was defined as the interval from PACU admission to discharge. (Page 9, Line 170-172)

PACU observation time and total PCU time were included as secondary outcomes and are now reported in the methods section of the abstract. (Page 2, Line 36-37)

-I can see from the main methods section that there are a lot of secondary outcomes and maybe it is not possible to list all in the abstract. But the authors should at least ensure that any outcomes reported in the abstract results are first introduced in the abstract methods.

Response: Thank you for your suggestion. We have added the relevant secondary outcomes in the methods section of the abstract. (Page 2, Line 35-38)

-I also suggest the authors select a single phrasing for each outcome and check the manuscript to ensure it is named consistently throughout, and ensure all outcomes are named in the relevant positions.

Response: Thank you for your comments. We have reviewed the manuscript to ensure that all results are clearly reported in the relevant sections.

Introduction

6. Clear and meets STROBE checklist criteria – no suggestions.

Response: Thank you very much.

Methods

7. Line 96. Undergoing should be “underwent”, or even better, “had”.

Response: Thank you for your careful review and valuable feedback. We have changed the word “undergoing” to “had”. (Page 6, Line 101)

8. Line 109. This sentence has unclear grammar and needs revising.

Response: Thank you for pointing out our mistake. We have revised the sentence as follows: All patients underwent preoperative fasting. (Page7, Line 115)

9. Line 153. The primary outcome is anesthesia recovery time (labelled T1), which is the cessation of anesthesia to extubation. The units of the outcome should be stated.

Response: Thank you for your suggestion. We have added the unit of minutes. (Page 9, Line 160-161)

10. Line 175-176. I am not sure a sample size calculation is needed given that this is a retrospective analysis, and the authors cannot control the number of participants available. However, if the authors wish to report this, please add the expected effect size, variance, and the references from which this was taken.

Response: Thank you for your comments. We have revised the manuscript. (Page 11, Line 200-205)

Based on our previous results, the mean extubation time for both groups were 19.00±11.52 min and 22.15±14.42 min, respectively. A two-tailed test with α set at 0.05, 90% power and a sample size of 1:1 indicated that a minimum sample size of 361

participants per group was required. As PSM will be used for case selection, we included a larger sample size to ensure that the final number after PSM met the required threshold.

11. Propensity score matching: I suggest moving this to separate subheading prior to the statistical analysis, because it is essentially a preparatory activity. There are also some details missing;

Response: Thank you for your suggestion. We have moved the section on propensity score matching to a separate subheading with prior to the statistical analysis, and added some missing details. (Page 10, Line 180-196)

-The rationale for using propensity score matching rather than standard regression (perhaps could be added to the introduction)

Response: Yes, the rationale for using propensity score matching rather than standard regression. However, due to potential differences in the distribution of covariates between the two groups, propensity score matching (PSM) is an effective method to adjust for such imbalances. We have included this explanation in the introduction: "Propensity score matching (PSM) was used to adjust pairs of patients with and without esketamine for potential confounders". (Page 6, Line 90-91)

-How were the matching variables chosen?

-Certain matching variables are ambiguous, e.g. surgical category (what categories), surgery duration (categorical or continuous?), medical history, please add details so these variables are better understood.

Response: Based on relevant studies and our clinical experience, we performed matching for variables in the baseline characteristics. Matching variables included age, gender, BMI, chronic disease, ASA physical status classification, surgical category, duration of surgery, duration of anaesthesia, intraoperative blood loss, and use of PCA, as described in the Methods section. (Page 10, Line 192-194)

Surgical categories included hepatobiliary surgery, gastrointestinal surgery and colorectal surgery. The duration of surgery was treated as a continuous variable (measured in minutes). The variable 'medical history' was revised to 'chronic disease', which included hypertension, diabetes, coronary heart disease and chronic obstructive pulmonary disease. Detailed information is shown in Table 1.

-It is stated that linear regression was used to estimate the matching, but typically PSM would use logistic regression to match receiving treatment 1 vs receiving treatment 2. R MatchIt also defaults to logistic regression. Could this be explained more – including all parameters specified in the MatchIt functions?

Response: Thank you for your comments. In this study, PSM was performed using logistic regression. Linear regression was then used to assess the association between study variables and outcomes. The results of the multiple linear regression analysis are shown in Supplementary Table 1.

Supplement Table 1. Results of multiple linear regression analysis

Variables	b	S.E	t	β (95%CI)	P	m_b	m_S.E	m_t	aβ (95%CI)	aP
Age	0.27	0.04	7.09	0.27 (0.20 ~ 0.35)	<.001	0.20	0.04	4.23	0.20 (0.12 ~ 0.27)	<.001
Gender										
Male				0.00 (Reference)						
Female	1.38	1.12	1.23	1.38 (-0.82 ~ 3.58)	0.218					
BMI	-0.89	0.17	-5.34	-0.89 (-1.21 ~ -0.56)	<.001	-0.60	0.16	-3.80	-0.60 (-0.91 ~ -0.30)	<.001
Chronic disease										
Hypertension										
No				0.00 (Reference)						
Yes	0.20	1.46	0.14	0.20 (-2.65 ~ 3.06)	0.888					
Diabetes										
No				0.00 (Reference)						
Yes	-1.78	2.41	-0.74	-1.78 (-6.51 ~ 2.94)	0.459					
Coronary heart disease										
No				0.00 (Reference)						
Yes	3.28	4.35	0.75	3.28 (-5.25 ~ 11.81)	0.451					
COPD										
No				0.00 (Reference)					0.00 (Reference)	
Yes	11.59	4.85	2.39	11.59 (2.10 ~ 21.09)	0.017	5.42	4.48	1.21	5.42 (-3.36 ~ 14.20)	0.226
ASA physical status										
I				0.00 (Reference)						
II	3.82	2.05	1.86	3.82 (-0.20 ~ 7.83)	0.062					
III	3.34	2.88	1.16	3.34 (-2.31 ~ 8.99)	0.246					
Surgery type										
Hepatobiliary surgery				0.00 (Reference)					0.00 (Reference)	
Gastrointestinal surgery	-3.97	1.29	-3.08	-3.97 (-6.50 ~ -1.44)	0.002	-3.70	1.23	-3.00	-3.70 (-6.12 ~ -1.28)	0.003
Colorectal surgery	7.89	1.39	5.68	7.89 (5.17 ~ 10.61)	<.001	1.20	1.60	0.75	1.20 (-1.92 ~ 4.33)	0.451
Surgery duration (min)	0.09	0.01	9.58	0.09 (0.07 ~ 0.11)	<.001	0.05	0.07	0.81	0.05 (-0.08 ~ 0.18)	0.417

Anesthesia duration (min)	0.09	0.01	9.80	0.09 (0.07 ~ 0.10)	<.001	-0.05	0.06	-0.77	-0.05 (-0.17 ~ 0.07)	0.442
Intraoperative blood loss (ml)	0.10	0.01	10.50	0.10 (0.08 ~ 0.12)	<.001	0.06	0.01	5.37	0.06 (0.04 ~ 0.08)	<.001
Esketamine										
No				0.00 (Reference)					0.00 (Reference)	
Yes	-3.35	1.11	-3.01	-3.35 (-5.53 ~ -1.17)	0.003	-2.83	1.02		-2.83 (-4.84 ~ -0.83)	0.006
PCA										
No				0.00 (Reference)					0.00 (Reference)	
Yes	12.55	1.13	11.06	12.55 (10.32 ~ 14.77)	<.001	6.04	1.67		6.04 (2.77 ~ 9.31)	<.001

ASA: American Society of Anesthesiologists; BMI: body mass index; COPD: chronic obstructive pulmonary disease; PCA: Patient controlled analgesia; PSM: propensity score matching

-How was the matching evaluated, and were any diagnostics/sensitivity analyses performed?

Response: It is generally accepted that a standardised mean difference (SMD) of less than 0.1 for all variables indicates a good fit [1]. We have chosen a value of 0.05. (Page 11, Line 184-192)

In addition, we didn't perform sensitivity analyses, and the lack of sensitivity analysis may affect the robustness of the results. This limitation is acknowledged in the "Strengths and limitations of this study" (Page 4) and in the limitations section of the manuscript. (Page 17, Line 335-338)

Results

12. Line 207-208. This approach to evaluating the PSM should be explained in the methods.

Response: Thank you for your kind advice. We have explained in the methods: "It is generally accepted that a standardised mean difference (SMD) of less than 0.1 for all variables indicates a good fit. We have chosen a value of 0.05." (Page 10, Line 184-195)

13. Figure 2. How was SMD calculated for categorical variables? This should be explained in the methods.

Response: Thank you for pointing that out. In propensity score matching (PSM), the standardised mean difference (SMD) for categorical variables is calculated as follows: For binary variables, the SMD is the difference in event rates between the treatment and control groups divided by the pooled standard deviation, where the pooled p is the weighted average of the event rates in both groups. For multi-category variables, the variable is split into several binary dummy variables and the SMD is calculated separately for each dummy variable, taking the maximum absolute value. Typically, an SMD ≤ 0.1 indicates good balance and an SMD ≤ 0.2 is considered acceptable. In the study provided by the user, the SMD for all categorical variables after adjustment was < 0.1, indicating a highly balanced distribution of covariates. We have added this in the methods section. (Page 10, Line 184-190)

14. Figure 3. Distance should be defined somewhere.

Response: Thank you for your suggestion. We have revised the Figure 3.

15. Table 1. The categories for the matching variables should be explained in the methods. The descriptive statistics are not explained – suggest adding to the table label. E.g. 58.00

(45.00, 66.00) – median and IQR? I don't think z-statistics or chi-square statistics are necessary to report because the reader cannot interpret these – P and SMD should be enough. Removing those statistics might solve some of the line-wrapping issues.

Response: Thank you for pointing that out. Descriptive statistics, including median and interquartile range (IQR), have been added in the table footnote. Following your suggestion, 'statistics' has been removed, and only P values and SMD are retained.

16. Line 215. The p-value threshold will be clearer if added to line 212 after "differences"

Response: Thank you for your suggestion. We have moved the p-value threshold as per your suggestion. (Page 12, Line 228)

17. Table 2. Data are presented as median/IQR or n/% - but which rows? I don't think "statistic" is needed? I think for full clarity, T2 and T3 should also be reported in the main text, to avoid reporting only significant results.

Response: Thank you for your comments. We have clearly indicated the use of median/IQR or n/% and removed 'statistic'. In addition, T2 and T3 have also been reported in the main text (Page 13, Line 258-263).

18. Table 3. Generally for all the tables I recommend reducing the line spacing to improve readability, and checking the alignment of values down columns.

Response: Thank you for your kind advice. We have reduced the line spacing and checked the alignment.

19. The primary assumption of PSM is that there is no unmeasured confounding. Have the authors considered performing sensitivity analyses such as estimating E-values to assess this? The discussion also should consider this qualitatively, currently the discussion does not mention confounding at all.

Response: Yes. Although the two groups were matched on several demographic factors, there remains the potential for residual confounding due to unmeasured variables affecting the propensity score analysis. In addition, we didn't perform sensitivity analyses, and the lack of sensitivity analysis may affect the robustness of the results. This limitation is acknowledged in the "Strengths and limitations of this study" (Page 4) and in the limitations section of the manuscript. (Page 17, Line 335-338)

Discussion

20. I am unsure why the study being single center is a limitation – the reason should be explained, plus the impact on the results.

Response: Thank you for your suggestion. The current study is its retrospective and being conducted at a single center, this may limit the generalisability of study results. We have mentioned in the limitation. (Page 17, Line 334-335)

Reference

1. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med.* 2009;28(25):3083-3107.

VERSION 2 - REVIEW

Reviewer	**3**
Name	**Dack, Kyle**
Affiliation	**University of Bristol, MRC Integrative Epidemiology Unit**
Date	**22-Apr-2025**
COI	

The authors have addressed most sugggestions I had to improve the statistical clarity. There are a few minor queries below, but overall the quality of the analysis and reporting is high.

Methods

"Based on our previous results, the mean extubation time for both groups were 19.00±11.52 min and 22.15±14.42 min, respectively." There should be a reference for this, assuming it is from a prior study, or some other form of detail to explain where this information was obtained.

"It is generally accepted that a standardised mean difference (SMD) of less than 0.1 for all variables indicates a good fit." This is reasonable but I suggest adding a reference to support the claim that it is generally accepted.

P10 lines 192-195. This section explains what the matching variables were, but does not explain the criteria for selecting them. It is implied based on other sections that this is to minimize risk of confounding, but I recommend the paper directly states this, e.g. "matching variables were selected based on prior literature where there was evidence of being potential confounders", because this is likely to be checked for in any future systematic reviews which include this paper.

VERSION 2 - AUTHOR RESPONSE

Reviewer: 3

Comments to the Author:

The authors have addressed most sugggestions I had to improve the statistical clarity. There are a few minor queries below, but overall the quality of the analysis and reporting is high.

Methods

"Based on our previous results, the mean extubation time for both groups were 19.00±11.52 min and 22.15±14.42 min, respectively." There should be a reference for this, assuming it is from a prior study, or some other form of detail to explain where this information was obtained.

Response: Thank you for your suggestion. We performed preliminary experiments with 20 patients in each group and recorded the extubation time for both groups, which were 19.00±11.52 min and 22.15±14.42 min, respectively. We have added additional explanations in the Methods. (Page 11, Line 201-203)

"It is generally accepted that a standardised mean difference (SMD) of less than 0.1 for all variables indicates a good fit." This is reasonable but I suggest adding a reference to support the claim that it is generally accepted.

Response: Thank you very much for your kind advice. We have added a reference [1] to the manuscript. (Page 10, Line 186, reference 20 in the manuscript)

P10 lines 192-195. This section explains what the matching variables were, but does not explain the criteria for selecting them. It is implied based on other sections that this is to minimize risk of confounding, but I recommend the paper directly states this, e.g. "matching variables were selected based on prior literature where there was evidence of being potential confounders", because this is likely to be checked for in any future systematic reviews which include this paper.

Response: Thank you for pointing this out. We have revised in the Methods and added a reference [2] to the manuscript. (Page 10, Line 192-193, reference 21 in the manuscript)

Reference

1. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment

groups in propensity-score matched samples. *Stat Med.* 2009;28(25):3083-3107.

2. Hasselager RP, Hallas J, Gogenur I. Inhalation or total intravenous anaesthesia and recurrence after colorectal cancer surgery: a propensity score matched Danish registry-based study. *Br J Anaesth* 2021;126(5):921-30.