Peer Review File

Article information: https://dx.doi.org/10.21037/atm-21-1612

Reviewer A

Comments to the Author:

This study evaluated the relative effects of 3 type surgical techniques after hypertensive BG ICH using a single-center observational study. After IPTW adjusted logistic regression, endoscopic evacuation seems to be superior to other two techniques. However, there is major big differences between the three groups, especially the patients who underwent open craniotomy had initial severe symptoms such as low GCS, more large ICH volume, and the most unfavorable prognosis. I am wondering that how many patients were matched in these analyses. In addition, several standard difference values in some analyses are > 0.2. First of all, the resolution of sup. Fig 1 is poor and not recognizable.

Minor comments 1. Why do authors deal with a categorical variable as age? Continuous variable in terms of age has more informative.

Comment 1: However, there is major big differences between the three groups, especially the patients who undaerwent open craniotomy had initial severe symptoms such as low GCS, more large ICH volume, and the most unfavorable prognosis. I am wondering that how many patients were matched in these analyses.

Reply 1: Thank you for this question. It was an observational cohort study and randomization was not used, therefore some baseline factors (e.g., GCS score, Hematoma volume, etc.) were not balanced among groups. For this reason, we applied inverse probability of treatment weighting (IPTW), which was a widely used method in non-randomized study, to adjust these baseline factors, but not propensity score matching (PSM). IPTW is different from PSM, PSM will exclude those subjects who are not matched in the effect analysis stage, however, IPTW uses weighting to balance the baseline factors. As thus, there are no subjects were excluded in our study and all of the 703 cases were included in the final effect analysis. More details about IPTW can be found in *Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. STAT MED. [Journal Article; Research Support, N.I.H., Extramural; Research Support, Non-U.S. Gov't; Review; Systematic Review]. 2015 2015-12-10;34(28):3661-79.*

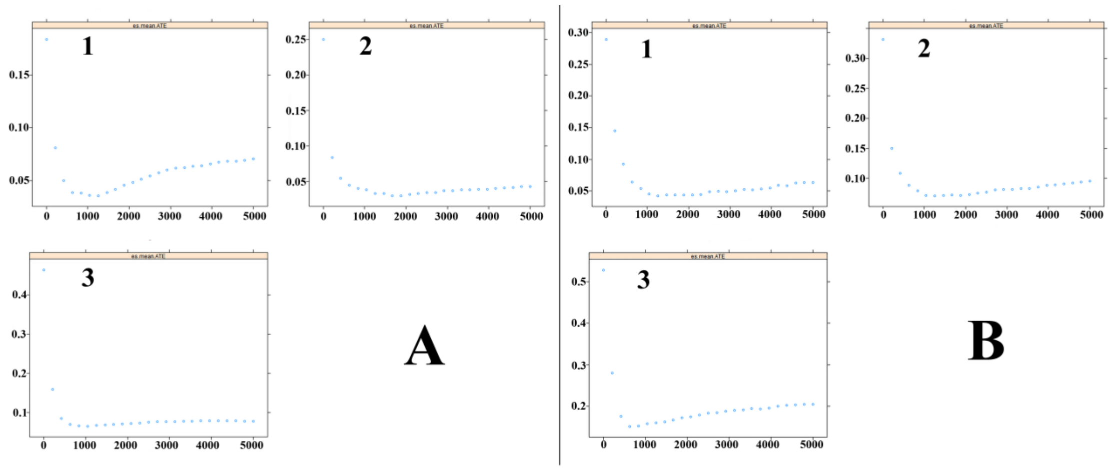
Changes in the text: we have added the "all the enrolled cases were included in the final effect analysis" to the **Statistical analysis** section in the revised manuscript (Page 11, lines 225-226).

Comment 2: In addition, several standard difference values in some analyses are > 0.2.

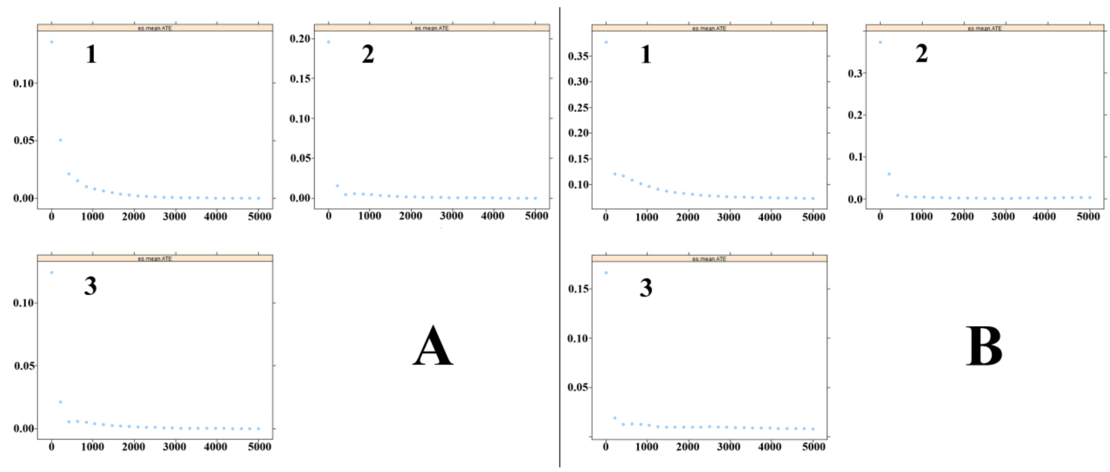
Reply 2: Thank you very much for your in-depth review of our manuscript. The major objective of IPTW is to balance the propensity score, which is an indicator to measure the extent of the unbalance of all the covariates as a whole. Therefore, IPTW can not ensure the balance of each covariate after weighted, and there may be some individual standard difference values in some analyses model are >0.2(1, 2).

What's more, the results showed that the standard mean difference of propensity score was <0.2 in each analysis and most of the standard mean differences of the covariates decreased a lot, so we believe that IPTW decreases the bias from the unbalanced covariates and increase the robustness of the estimation of group effects(3). We are sorry that this part may not clear in the original manuscript. In addition, we added the results of generalized boosted model to obtain the optimal balance of the propensity score among groups in **Supplemental Fig 1-7** in the revised supplemental material. Two of those Supplemental Figures are listed below.

Changes in the text: we added the results of the optimal balance of the propensity score among groups in the revised supplemental material. (**Supplemental Fig 1-7**).



Supplemental Fig 1. For the entire cohort, generalized boosted model to obtain the optimal balance of the propensity score among groups of the inverse probability of treatment weighted (IPTW) model associated with modified Rankin Scale (mRS) score (A) and mortality rates (B). Horizontal axis represents the iteration times and vertical axis represents the balance measure. (1) Balance for EE against others. (2) Balance for SA against others. (3) Balance for OC against others. SA=Stereotactic Aspiration, EE=Endoscopic Evacuation, OC=Open Craniotomy.



Supplemental Fig 5. For the subgroup of GCS score 3–5, generalized boosted model to obtain the optimal balance of the propensity score among groups of the inverse probability of treatment weighted (IPTW) model associated with modified Rankin Scale (mRS) score (A) and mortality rates (B). Horizontal axis represents the iteration times and vertical axis represents the balance measure. (1) Balance for EE against others. (2) Balance for SA against others. (3) Balance for OC against others.

SA=Stereotactic Aspiration, EE=Endoscopic Evacuation, OC=Open Craniotomy.

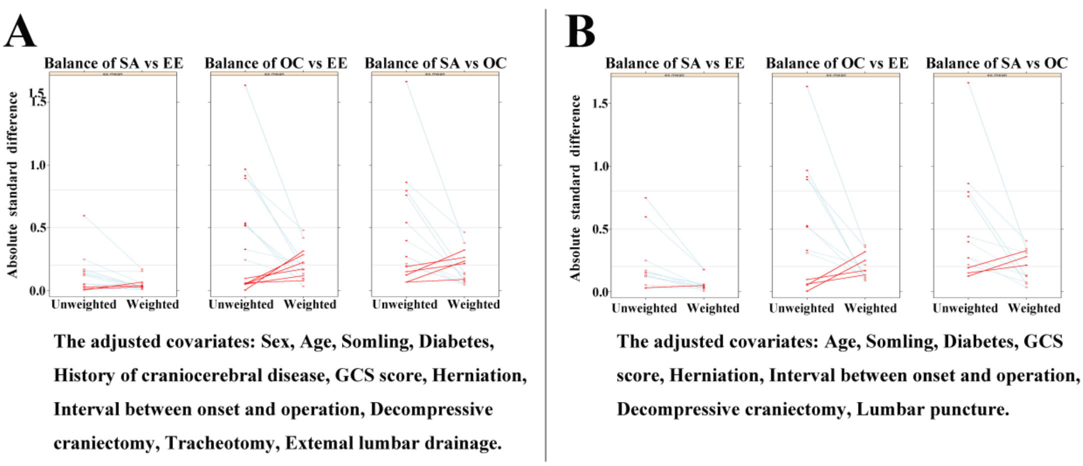
References:

1. McCaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. STAT MED. [Journal Article; Research Support, N.I.H., Extramural; Research Support, U.S. Gov't, P.H.S.]. 2013 2013-08-30;32(19):3388-414.
2. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. STAT MED. [Journal Article; Research Support, N.I.H., Extramural; Research Support, Non-U.S. Gov't; Review; Systematic Review]. 2015 2015-12-10;34(28):3661-79.
3. Ridgeway, G., McCaffrey, D. F., Morral, A. R., Burgette, L. F. & Beth Ann, G. Toolkit for Weighting and Analysis of Nonequivalent Groups: A Tutorial for the R TWANG Package, https://www.rand.org/pubs/tools/TL136z1.html (2014)

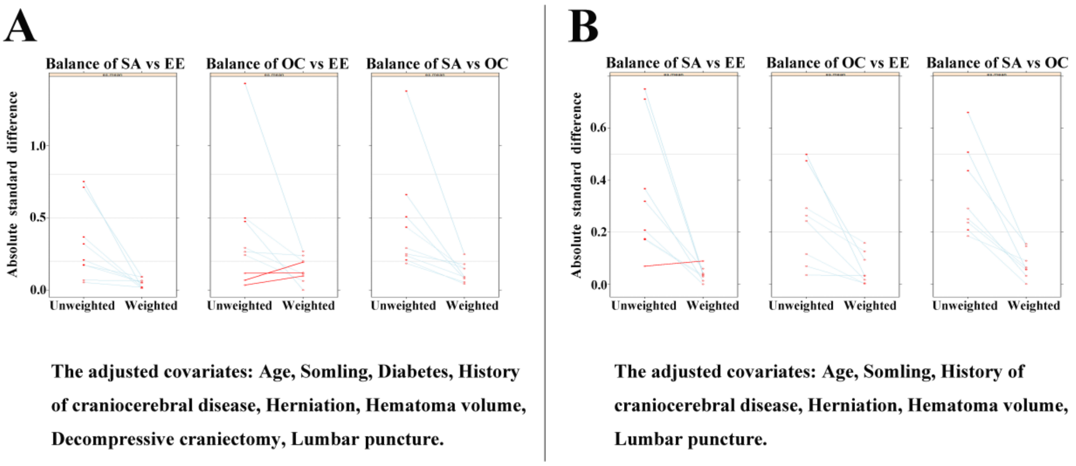
Comment 3: the resolution of sup. Fig 1 is poor and not recognizable.

Reply 3: We apologize for the poor quality of sup. Fig 1. The sup. Fig 1 has been divided into 7 figures with high quality (**Supplemental Fig 8-14**) in the revised supplemental material. Two of those Supplemental Figures are listed below.

Changes in the text: Supplemental Fig 8-14



Supplemental Fig 10. For the subgroup of hematoma volume of 40–80 mL, results of the standardized mean differences to evaluate the effect of weights on the magnitude of each confounding factors of the inverse probability of treatment weighted (IPTW) model associated with modified Rankin Scale (mRS) score (A) and mortality rates (B). SA=Stereotactic Aspiration, EE=Endoscopic Evacuation, OC=Open Craniotomy.



Supplemental Fig 13. For the subgroup of GCS score 6–8, results of the standardized mean differences to evaluate the effect of weights on the magnitude of each confounding factors of the inverse probability of treatment weighted (IPTW) model associated with modified Rankin Scale (mRS) score (A) and mortality rates (B). SA=Stereotactic Aspiration, EE=Endoscopic Evacuation, OC=Open Craniotomy.

Comment 4: Minor comments 1. Why do authors deal with a categorical variable as age? Continuous variable in terms of age has more informative.

Reply 4: Thank you for your comments. We agree with you that treating age as a continuous variable is more informative. However, the clinical meaning can not be guaranteed if the effect size was estimated as one year increase in age. Therefore, we classified the population into two old and young with the cutoff age of 60.

Reviewer: B

Comments to the Author:

The authors conducted a retrospective cohort analysis to evaluate the outcome and mortality in patients with basal ganglion hemorrhage who received different surgical procedures. The work showed a better outcome and reduced mortality rate in patients who received EE compared with patients who received SA. In addition, the difference can be further stratified by hematoma volume and GCS scores. Thus, the data can provide clinicians more information regarding the choice of surgical methods in these kinds of patients.

I have several comments:

1. The patient number may be wrong in the result section, especially the patient number who received OC.
2. The author should define what is cerebral disease in page 11, line 3.
3. The supplementary figure 1 is too blurred to see clearly.
4. There is no table 2 in the manuscript.
5. Multivariate logistic regression analysis may also adjust the impact of confounding factors on the primary outcome. Why do the authors further preform IPTW method to balance the baseline difference of covariate? The 6-M mortality rate is different between EE and OC group in different statistical methods. The authors may give some discussion or how to explain it.
6. What is the statistical method in subgroup analysis?
7. The authors had a previous article published in 2020, which used a smaller cohort and different statistical methods. The results showed a different conclusion in functional outcomes between EE and SA group. The authors may give discussion regarding it.
8. In the figure 1, the patient number is also wrong (241+68+703 is not equal to 1016)
9. The authors may give one section regarding how they make a decision regarding the surgical methods in different patient populations.

Comment 1: The patient number may be wrong in the result section, especially the patient number who received OC.

Reply 1: We thank the reviewer for his/her carefulness to point out the inconsistency of the number of patients in OC group. The wrong number “143” in the revised manuscript has been corrected to “148” in the **Result** section (Page 12, line 244).

Changes in the text: we modified the number to “148” in the **Result** section (Page 12, line 244).

Comment 2: The author should define what is cerebral disease in page 11, line 3.

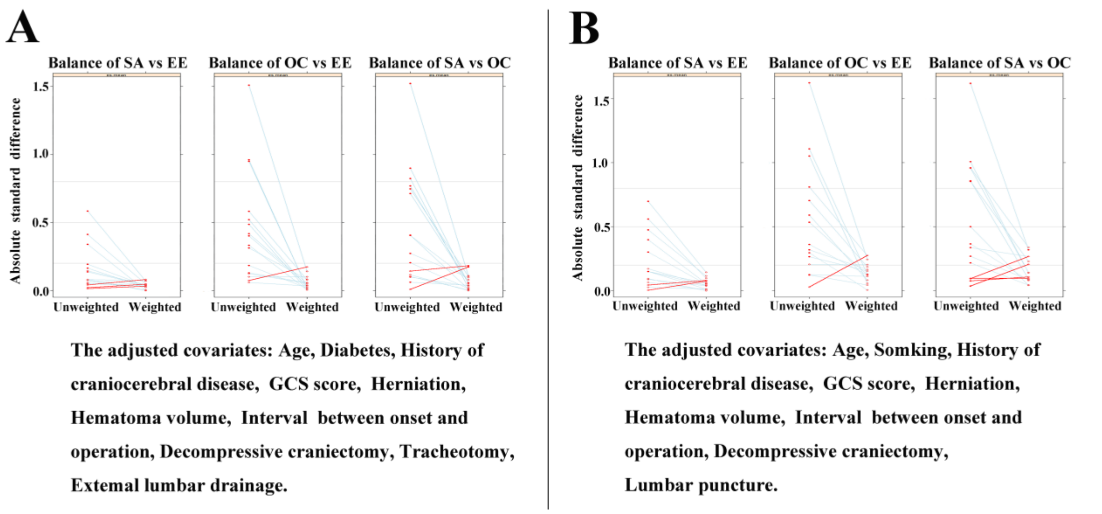
Reply 2: Thank you for the in-depth review and valuable suggestion. History of craniocerebral disease was defined as cerebral hemorrhage, cerebral infarction, and traumatic brain injury. We added the relevant definition to the **Data collection and outcomes evaluation** section (Page 9, lines 193-194) in the revised manuscript based on your suggestion. Thank you so much.

Changes in the text: we have added the definition of cerebral disease to our text as suggested (Page 9, lines 193-194).

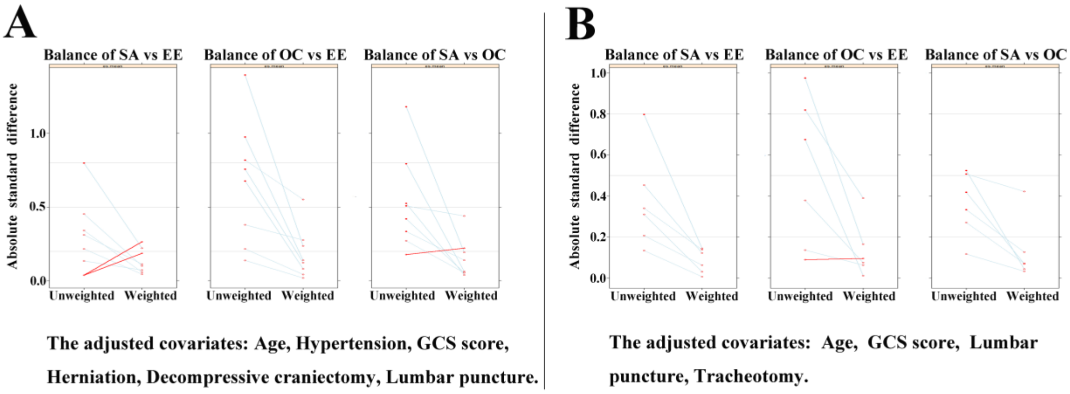
Comment 3: The supplementary figure 1 is too blurred to see clearly.

Reply 3: We apologize for the poor quality of supplementary figure 1. The supplementary figure 1 has been divided into 7 figures with high quality (**Supplemental Fig 8-14**) in the revised supplemental material. Two of those Supplemental Figures are listed below.

Changes in the text: Supplemental Fig 8-14



Supplemental Fig 8. For the entire cohort, results of the standardized mean differences to evaluate the effect of weights on the magnitude of each confounding factors of the inverse probability of treatment weighted (IPTW) model associated with modified Rankin Scale (mRS) score (A) and mortality rates (B). SA=Stereotactic Aspiration, EE=Endoscopic Evacuation, OC=Open Craniotomy.



Supplemental Fig 11. For the subgroup of hematoma volume of ≥80 mL, results of the standardized mean differences to evaluate the effect of weights on the magnitude of each confounding factors of the inverse probability of treatment weighted (IPTW) model associated with modified Rankin Scale (mRS) score (A) and mortality rates (B). SA=Stereotactic Aspiration, EE=Endoscopic Evacuation, OC=Open Craniotomy.

Comment 4: There is no table 2 in the manuscript.

Reply 4: Thank you for pointing out this important point. We did submit it at the beginning, but it was not shown in the review document for unknown reasons. We are so sorry about that. Table 2 is listed below for your convenience to review. We make sure that it is included in the revised manuscrip.

Table 2. The results of the multivariate logistic regression model and IPTW-adjusted logistic regression model exploring the 6-month mortality rates and a poor modified.Rankin Scale (mRS) score.

SA=Stereotactic Aspiration, EE=Endoscopic Evacuation, OC=Open Craniotomy.

Comment 5: Multivariate logistic regression analysis may also adjust the impact of confounding factors on the primary outcome. Why do the authors further preform IPTW method to balance the baseline difference of covariate? The 6-M mortality rate is different between EE and OC group in different statistical methods. The authors may give some discussion or how to explain it.

Reply 5: Thank you for the in-depth review and insightful comments. In the observational study, multivariate logistic regression analysis and propensity

Outcomes	SA (n=343)	EE (n=212)	OC (n=148)	Multivariate logistic regression OR (95% CI)		IPTW-adjusted analysis OR (95% CI)	
				SA VS EE	OC VS EE	SA VS EE	OC VS EE
mRS score (4-6)	181(52.77%)	94(44.34%)	117(79.05%)	1.746 (1.135-2.686)	0.768 (0.358-1.648)	1.359 (1.091-1.692)	0.956 (0.765-1.194)
Mortality rate	124(36.15%)	34(16.04%)	71(47.97%)	3.045 (1.831-5.062)	1.422 (0.781-2.588)	2.396 (1.865-3.080)	1.395 (1.059-1.837)

score analysis are commonly used statistical analysis methods to control confounding factors. In this present study, we used two methods simultaneously and verified each other.

For the multivariate logistic regression analysis, the more confounding factors are included, the more outcome events are needed. Usually, the number of outcome events should be at least 15-20 times that for covariates. On the contrary, the advantage of IPTW method is to replace multiple confounding factors with one propensity score value, reduces the number of independent variables. Therefore, overcome the limitation of the number of independent variables in multivariate logistic regression analysis, and effectively reduce the requirement of sample size. At the same time, in the IPTW model, except for the different distribution of exposure or treatment factors and outcome variables, it can be considered that other confounding factors between groups are balanced and comparable, which is equivalent to “post-randomization”, so that the data of observational studies can achieve the effect of approximately random distribution(1). Due to the limitations of our data sample size and complex confounding factors, IPTW method is more appropriate and the results are more stable(2). As indicated by our results, the 95% CI of IPTW method is smaller than that of multivariate logistic regression analysis.

Both two statistical methods showed an increased risk of 6-M mortality rate of the OC group. The IPTW method showed that the results were statistically significant (OR 1.395, 95%CI 1.059-1.837), but multivariate logistic regression analysis showed that the results were not statistically significant (OR 1.422, 95%CI 0.781-2.588). This may be caused by the characteristics of our data and the differences in the principles of the two statistical methods. A larger sample size may be needed to obtain significant results in the multivariate logistic regression analysis, while the IPTW method is more suitable for processing data with smaller sample size and complex confounding factors. Therefore, we are more inclined to the results of the IPTW method among two statistical methods.

References:

1. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. [Journal Article]. 2011 2011-05-01;46(3):399-424.

2. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. STAT MED. [Journal Article; Research Support, N.I.H., Extramural; Research Support, Non-U.S. Gov't; Review; Systematic Review]. 2015 2015-12-10;34(28):3661-79.

Comment 6: What is the statistical method in subgroup analysis?

Reply 6: Thank you for this question. We are sorry that this part may not clear in the original manuscript. The statistical method in subgroup analysis is also the IPTW-adjusted logistic regression analysis. The variables used for grouping was not included in the subgroup analysis model. We have added the “In the subgroup analysis, a similar IPTW-adjusted logistic regression analysis was adopted” and “the grouping factor of the subgroup analysis was not included in the IPTW model" to the **Statistical analysis** section in the revised manuscript (Page 11, lines 226-227,229).

Changes in the text: we have modified our text as advised to make it clearer (Page 11, line 226-227).

Comment 7: The authors had a previous article published in 2020, which used a smaller cohort and different statistical methods. The results showed a different conclusion in functional outcomes between EE and SA group. The authors may give discussion regarding it.

Reply 7: Thank you very much for your thorough reading of our manuscript and the insightful comment. The conclusion of the previous article was EE can decrease the 6-month mortality of spontaneous basal ganglia hemorrhage, especially in patients with a hematoma volume of ≥40 ml. In previous article, we found a difference in mortality between EE and SA groups, but there was no significant difference in the functional outcome. In terms of functional outcomes, the mRS score of 0-2 was regarded as good and 3-5 was regarded as poor in previous article. However, the proportion of mRS=3 in SA, EE, and OC groups of the entire cohort in this present study was 20.4%, 32.5%, and 18.2%, respectively. The proportion of mRS=3 in the EE group was much higher than that in the other two surgical groups.

In some studies, mRS=3 was attributed to poor functional outcomes(3, 4), while in others, mRS=3 was attributed to good functional outcomes(5, 6). In the recently published MISTIE III trial(7), mRS=3 was attributed to good functional outcomes. This setting was referred to in our study. Good functional outcome was defined as the patients who achieved an mRS score of 0-3. The poor functional outcome was defined as mRS score of 4-6. We found that when the hematoma volume of patients was ≥ 40ml, EE can decrease mortality and improve functional outcomes compared with SA.

Accordingly, we have added relevant discussions to the **DISCUSSION** section in the revised manuscript (Page 17, lines 353-357).

Changes in the text: we have modified our text regarding the relevant discussions (Page 17, lines 353-357).

References:

3. Wang WZ, Jiang B, Liu HM, Li D, Lu CZ, Zhao YD, et al. Minimally invasive craniopuncture therapy vs. conservative treatment for spontaneous intracerebral hemorrhage: results from a randomized clinical trial in China. INT J STROKE. [Journal Article; Multicenter Study; Randomized Controlled Trial]. 2009 2009-02-01;4(1):11-6.

4. Choo YS, Chung J, Joo JY, Kim YB, Hong CK. Borderline basal ganglia hemorrhage volume: patient selection for good clinical outcome after stereotactic catheter drainage. J NEUROSURG. [Journal Article]. 2016 2016-11-01;125(5):1242-8.

5. Vespa P, Hanley D, Betz J, Hoffer A, Engh J, Carter R, et al. ICES (Intraoperative Stereotactic Computed Tomography-Guided Endoscopic Surgery) for Brain Hemorrhage: A Multicenter Randomized Controlled Trial. STROKE. [Journal Article; Multicenter Study; Randomized Controlled Trial]. 2016 2016-11-01;47(11):2749-55.

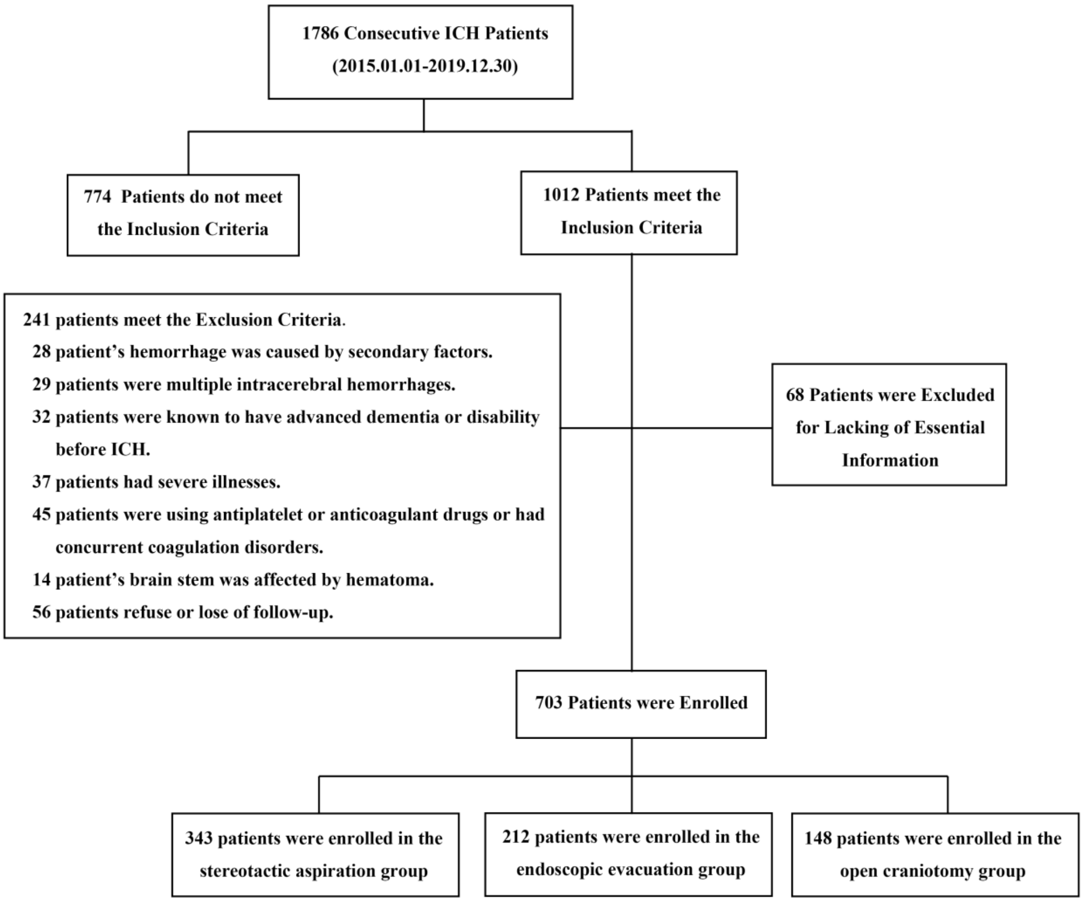
6. Hanley DF, Lane K, McBee N, Ziai W, Tuhrim S, Lees KR, et al. Thrombolytic removal of intraventricular haemorrhage in treatment of severe stroke: results of the randomised, multicentre, multiregion, placebo-controlled CLEAR III trial. LANCET. [Clinical Trial, Phase III; Journal Article; Multicenter Study; Randomized Controlled Trial]. 2017 2017-02-11;389(10069):603-11.

7. Hanley DF, Thompson RE, Rosenblum M, Yenokyan G, Lane K, McBee N, et al. Efficacy and safety of minimally invasive surgery with thrombolysis in intracerebral haemorrhage evacuation (MISTIE III): a randomised, controlled, open-label, blinded endpoint phase 3 trial. LANCET. [Clinical Trial, Phase III; Journal Article; Multicenter Study; Randomized Controlled Trial]. 2019 2019-03-09;393(10175):1021-32.

Comment 8: In the figure 1, the patient number is also wrong (241+68+703 is not equal to 1016)

Reply 8: Thank you for the careful review of our manuscript and pointing out this important point. We apologize for this error. We worked on the manuscript for a long time and the repeated addition and reduction of patients according to the inclusion and exclusion criteria may lead to this mistake. The patient number has been carefully checked and corrected in **Figure 1** and **Results** section in the revised manuscript. we thank again the reviewer for your in-depth review to improving the quality of this manuscript.

Changes in the text: we have modified the wrong patient number in **Figure 1** and **Results** section (Fig.1 and Page 11, lines 238-239). **Figure 1** is listed below for your convenience to review.



Comment 9: The authors may give one section regarding how they make a decision regarding the surgical methods in different patient populations.

Reply 9: Thank you for your valuable suggestion. In the clinical treatment process, surgeons emergent selected the surgical technique according to the volume, location, and progression of the hematoma, the patient's general condition, the patient's family's wishes, and the surgeons' experience. For patients with larger hematoma volume or lower GCS scores, the probability of intracranial hypertension or cerebral herniation is higher during the operation, surgeons might more quickly manage the emergent intraoperative intracranial hypertension or cerebral herniation when choosing OC.

Accordingly, we have added the relevant section to the revised manuscript (Page 8, lines 158-163).

Changes in the text: we have modified our text in the revised manuscript as suggested (Page 8, lines 158-163).