Peer review file

Article information: https://dx.doi.org/10.21037/jtd-21-1063

Reviewer A

This paper is a retrospective study to investigate the learning curve of minimally invasive esophagectomy (MIE) with comparing between two different situations. In hospital 1, an experienced surgeon performed MIE in new circumstance, and in hospital 2, the surgeons who had experienced sufficient cases of open esophagectomy and minimally invasive surrogate surgery started and performed MIE. The authors suggested that learning phase of the individual surgeon was more important for the outcomes of MIE than the institutional learning phase. Furthermore, they demonstrated that high-level of surrogate surgery might shorten the learning curve.

Comment 1: However, the number of surgeons who participated in this program in hospital 2 was not mentioned and they compared the outcomes of only 2 institutions. Moreover, the anastomotic procedure was changed during this study period. The procedure switch of anastomosis extremely influences the outcome of anastomotic leakage which is one of the primary endpoints of this study.

Authors response: We appreciate the comment from the Reviewer. In Hospital 2, the team included two gastrointestinal surgeons and a thoracic surgeon. The team remained the same during this learning curve. A small addition was made in Methods section.

The Reviewer points out the fact that anastomosis technique was changed during the study period. This probably has effect on leak rates (rate decreased over time). However, we consider this as an important part of learning curve, since during learning curve and mastering of complex technique, small refinements are made, otherwise there would be no improvements.

Please also see our responses to other Reviewers. We have updated the study cohort which now includes 132 patients from Hospital 1 and 57 patients from Hospital 2.

Comment 2: I don’t think that the conclusions of this study could be clearly reached by their results.

Authors response: In conclusion, we state that “…learning phase of an individual surgeon determines the outcomes of MIE, not the institutional volume or learning phase” We assume that the Reviewer agrees with that cited text based on our results. However, we agree that the following “Compared to that presented in the literature, high-level of surrogate surgery seems to shorten the learning curve” is not entirely supported by our results. Also, please see comments from Reviewer 3 and our response. In the abstract and conclusions sections the final sentence was omitted.

Following change was made, Abstract and Conclusions section: The following sentence was omitted “Compared to that presented in the literature, high-level of surrogate surgery seems to shorten the learning curve.”

Minor revision
P9. L207; Figure 1c→Figure 2e
P9. L207; Figure 2c→Figure 2f

Authors response: We appreciate Reviewer’s accuracy. Changed were made accordingly.

Reviewer B

Many thanks for the oppertunity to review this manuscript. I read it with interest. This study adds to the understanding of learning curves by reporting a learning curve from a surgeon who got proficient and switched clinics vs a surgeon who started learning from scratch. I don't think this type of learning curve comparison has been published previously?
Please find below some points that are meant to improve the manuscript.

Authors response: We thank the Reviewer for these overall positive comments. Please also see our responses to other Reviewers. We have updated the study cohort which

now includes 132 patients from Hospital 1 and 57 patients from Hospital 2.

Abstract:

Comment 3: Please define primary outcome parameters. How was AL diagnosed for example (as the ECCG doesn't provide guidance for this)?

Authors response: We added outcome parameters to Abstract and to Background section. Individual primary and secondary outcome parameters have been listed in Abstract lines 52-53.

Also, AL diagnostic method was added to Methods section lines 100-101.

Following changes were made: Abstract

“The aim was to evaluate the learning curve for MIE focusing on short-term outcomes in two settings”

“The primary outcomes were major complications and anastomosis leaks. Secondary outcomes were operative time, blood loss, lymph node yield, hospital stay and 1-year mortality.”

Methods section

“…anastomotic leak rate (diagnosed with either endoscopy or computed tomography (CT) with oral contrast)”

Comment 4: I'm not sure about the intention-to-treat statement in the abstract. For example, if a patient was metastasized at operation, you still included this patient in analysis? I don't think this should be done.

Authors response: By intent-to-treat basis we mean that conversions were included in the study. Patients with occult metastases were not included. We clarified this in Methods sections.

Following changes were made: “Benign diseases and operations aborted due to

inoperable disease, ie. occult metastases or carcinosis, were excluded."

Methods:

Comment 5: Please elaborate on the previous experience of the surgeons: how many cases of open and MIO were done before this study? How did surgeon 1 learn MIO? (teach himself? fellowship at another clinic?) Surgeon age? etc.

Authors response: This data was clarified in the Methods section. The surgeon in Hospital 1 performed first hybrid operations in Helsinki University hospital in 2007. In 2009, after his colleague had finished his fellowship at Division of Thoracic and Foregut Surgery at University of Pittsburgh, they started the formal MIE program. After performing 49 MIEs or hybrid operations and more than 100 open eophagectomies, surgeon at the age of 44 started the program at the Central Finland Central Hospital.

Following changes were made: In Hospital 1, MIE program started in September 2012 by a surgeon who had done his first self-learned hybrid esophagectomy in 2007. In 2009, after his colleague finished MIE fellowship, they started formal MIE program in Helsinki University Hospital. After performing 49 MIEs or hybrid procedures and more than 100 open esophagectomies together with a significant experience of minimally invasive lung cancer surgery, new program was started (11).

Comment 6: What was the yearly volume of the hospitals?

Authors response: Annual volume in Hospital 1 during the study period was 15 MIEs and in Hospital 2 it was 14 MIEs. This was added in the Results section.

Following changes were made: During the study, the mean annual MIE volume for cancer was 15 in Hospital 1 and 14 in Hospital 2.

Comment 7: Could you specify what type of neoadjuvant treatment the patients had?

Authors response: The given neoadjuvant treatment was strictly guidelines based. Reference to ESMO guidelines is provided in the Methods section. Also we added

reference to our previous study where neoadjuvant treatment regimen has been described in detail.

Following changes were made: Neoadjuvant treatment regimen has been previously described in detail (11).

Comment 8: Please define learning associated morbidity: how did you calculate this?

Authors response: We thank the Reviewer for accuracy. Actually, we have used the term "learning associated morbidity" if learning curve in complication rates was seen and therefore patients in the early period experienced more complications due to learning. We agree with the Reviewer that learning associated morbidity should probably be used only when calculating some exact numbers or rates of excess complications due to learning phase. We have now omitted "learning associated morbidity" from the paper and replaced it with more representative terms such as "higher morbidity during the early learning curve.

Results:

Comment 9: Were there any Orringer or McKeown type MIO's done in the same time period?

Authors response: Yes, previously we stated only the number of Ivor Lewis procedures in the Methods section. The rest of the operations were McKeown type MIEs. The sentence is now in the following form: "A minimally invasive Ivor Lewis esophagectomy with en-bloc lymphadenectomy and intrathoracic anastomosis was the preferred approach in both Hospital 1 (n=116, 87.9%) (11) and Hospital 2 (n=55, 96.5%). The rest of the operations were McKeown type MIEs."

Comment 10: Figure regarding LN yield: is this a mirrorred image? In contrast to complications etc more LN harvest is better and therefore I would expect the CUSUM curve to be inversed? If you inversed it for interpretation purposes please state this clearly. If not: why did LN yield decrease towards the end in hospital 1 (although differences are small)

Authors response: Again, we thank for valuable comments. As the Reviewer presumes, LN curve was inverted to achieve similar presentation in all CUSUM curves (ie. ascending curve suggesting non-completed learning curve, and after plateau events occurred less often than expected, and in case of LN curve, higher yield than expected after plateau). This way the reader can see similar meaning in all curves, which is descending curve is preferable. As the Reviewer suggest, Figure legend was revised to state this clearly.

Following changes were made: "Curve was inverted regarding lymph node yield where descending curve means higher than expected yield."

Comment 11: Regarding hospital 2: I think the case series is too short to adequately define how long the learning curve is. If the case series is limited, you can never find a long learning curve. This is even a bigger caveat if (RA)CUSUMS are used since they measure regression to average (works well with long plateau's only), instead of for example moving average. Please discuss this properly as limitation and don't overstate your conclusions from this cohort

Authors response: We agree with the Reviewer that due to the small number of operations in Hospital 2, it is possible that length of the learning curve can be underestimated and future studies with larger cohort, especially from Hospital 2, is needed to confirm the suggested cut-offs. We added some critical discussion in the limitation section. However, even though it is possible that true learning curve is longer, it is noteworthy that outcomes in both hospitals were mostly superior compared to previously published Benchmark values suggesting at least fairly mastered procedure. Conclusion was softened.

Additionally, we updated the study cohort which now includes 132 patients from Hospital 1 and 57 patients from Hospital 2 in order to provide more reliable results.

Following changes were made: With a small number of patients, definitive conclusions cannot be drawn. Especially in Hospital 2, the small cohort size can give false results

regarding the true length of the learning curve. With a small cohort, it is possible that the learning curve has not been completed and future studies with an increased number of patients are needed. With these results recently published Benchmark outcomes were, however, mostly reached suggesting at least nearly completed learning phase.

Comment 12: I think hospital 2 has not yet completed the learning curve judging from this data

Authors response: We agree with the Reviewer that even with updated cohort size, it is possible that learning curve in Hospital 2 is not yet completed. This question is highly related to the previous one, please see our response to Q11.

Discussion

Comment 13: Be careful with statements on learning associated morbidity as this has not been calculated in the analysis (or even better: calculate it!)

Authors response: Please see our answer to COMMENT8. The term learning associated morbidity was omitted from the text. With small number of events in primary outcomes (major complications and leaks) even with updated data, we consider it best not to include learning associated morbidity. However, median rates of other outcomes before and after learning occurred have been presented providing some estimation of excess treatment in early period.

Comment 14: Good discussion and translation to clinical implications

Authors response: We thank the Reviewer for accurate comments which significantly improved the quality of our manuscript.

Reviewer C

The presented results are commendable. The authors explored learning curves with the RA-CUSUM Method for MIE in two hospitals establishing a MIE-programme in two different settings. 174 patients were retrospectively included. 1. centre: experienced

MIE surgeon in new hospital, 2. centre: surgeons experienced with open esophagectomy and minimally invasive surrogate surgery. The concept of this manuscript was understandable which might be worth publishing in the JTD.

Authors response: We thank the Reviewer for these positive comments.

However, the detail raised some questions and requests listed below. Especially the small case number in hospital 2 with non-reliable conclusions might lead to a rejection when not improved.

Authors response: Below we provide point-to-point responses, but see also Reviewer B comments and our responses regarding small cohort size in Hospital 2. We revised conclusions and softened expressions regarding completion of learning curve throughout. Furthermore, we updated the study cohort which now includes 132 patients from Hospital 1 and 57 patients from Hospital 2.

Comment 15: line 102-109: could you please explain and clarify the fact, that in both hospitals only a median volume of 16 MIE per year was performed. That seems to be a quiet small volume, whereas you describe the hospitals as the only centres performing esophagectomies in their region (line 232-233).

Authors response: It is true that due to esophageal cancer the mean annual number of MIEs in Hospital 1 was 15 and in Hospital 2 14. Finland has a population of only 5.5 million. These two hospitals cover roughly 23% of this Finnish population. The mean annual number of esophagectomies in Finland in the most recent publication was only 80. Therefore, these 2 centers were active in esophageal cancer surgery performing 36% of annual national esophagectomies. Overall in 2014, 8 hospitals in Finland were performing esophagectomies with annual median volume of 8.5 per hospital.

Following changes were made: Results section

“During the study, mean annual MIE volume was 15 in Hospital 1 and 14 in Hospital

2."

Comment 16: Line 147-149: Could you please further describe the convergence problems and your solution.

Authors response: As the number of females and stage IV patients were both low, there were situations where there were no patients in that group with the outcome. When logistic regression is calculated, calculations with zero outcomes lead to the maximum likelihood approaching infinity. (Albert A, Anderson J. On the existence of maximum likelihood estimates in logistic regression models. Biometrika. 1984;71(1):1–10. doi: 10.1093/biomet/71.1.1.)

Practically, this leads to unstable regression estimates, with regression coefficients and confidence intervals approaching infinity. The number of iterations calculated (and the magnitude of results) vary greatly between statistical programs, and the results are then not reliable. The calculated individual risk of anastomotic leak in these (Stage IV or female) patients is approaching 100% or 0%, while that of the others is distorted. When applied to racusum charts, the curves become distorted and false.

As stage III and Stage IV are the most similar categories under tumor stage, and the recommended practice for nonconvergnce in categorical variables is to combine similar groups together, we combined stages III and IV together to avoid nonconvergence. For dichotomous variables, such combination cannot be done. As the effect of sex on anastomotic leaks was assumed to be minor, sex was left out from the regression model for anastomotic leak.

We revised the text in statistical analysis section, which is now in the following form regarding this convergence problem.

Following changes were made: Methods section
"No or very few outcomes occurred in some logistic regression models including stage IV or female patients as separate categories. As the real effects of these variables on

outcome risk were clinically assumed to be minor, tumor stage was grouped 0-I, II, or III-IV in analyses for anastomotic leak, major complications, length of stay, and 1-year mortality, and sex was excluded from the model for anastomotic leak to avoid nonconvergence of logistic regression."

Comment 17: Figure 1a: Could you please describe and explain the secondary increase in operative time for Hospital 1.

Authors response: We thank the Reviewer for this comment. We had no proper explanation for this and went through all patients case by case. Fortunately, we detected an error in the dataset due to update in operation software from case 82. OR time had replaced operative time. We have now corrected this error and a new CUSUM curve has been drawn.

Comment 18: Especially for Hospital 2 when regarding Figure 2 it must be doubted that the learning curve is completely depicted, due to the small case number (n=46). This point has to be extensively discussed. Form the available data I cannot be concluded, that the learning curve is shorter than described in the literature. Ideally some more patients can be included in the study.

Authors response: Please see also previous comments. Cohort was updated to include 132 patients from Hospital 1 and 57 patients from Hospital 2.

This is important critical comment, please see also Reviewer B comments and our responses regarding this matter. We revised the manuscript throughout to avoid misleading conclusions. However, the results were superior compared to Benchmark study from high-volume specialist centers which suggest actual completion of learning curve (with CUSUM method meaning events occurred less often than expected). We could ask how much better could / should the results be compared to Benchmark? Still, it is possible that Benchmark study does not actually present the "best possible" results due to learning curves of their own.

Overall, we think that in Hospital 2 the learning curve seems to be nearly completed

(based on CUSUM), but regarding methodological shortcomings and real possibility that in future cohort study from these two centers the learning curve can differ from currently presented, conclusions were revised in order to avoid faulty statements.