Peer Review File

Article information: https://dx.doi.org/10.21037/ales-23-65

Reviewer A

While I find the study interesting, there are three major points that I believe require attention and revision for the article to be suitable for publication in our journal. I would like to highlight these points for your consideration and revision. Your efforts in addressing these concerns would be crucial in enhancing the overall quality and impact of the manuscript. I look forward to seeing the revised version of your article.

1st. While the study is informative, it is important to address the lack of impact due to limited introduction of new knowledge. Given the existing body of literature on thoracoscopic compared to open thoracotomy, it would be beneficial to highlight the unique contributions and strengths of your paper that distinguish it from previous reports.

The literature on the subject is from large high-volume centres or multi-centre studies. Such data may factor selection bias due to strict inclusion criteria and may also not factor the difficulties in setting up a new service as well as the operators' learning curve. Hence such data may not be directly translatable to lower volume or non-academic units. Our audit demonstrates not only demonstrates the transition and outcomes in a lower volume unit but also seeks to disseminate the criteria/benchmarks necessary for appropriate clinical governance when introducing this surgical service (introduction last paragraph and discussion first two paragraphs in yellow).

2nd. Identifying and emphasizing the novel aspects of your study will enhance the significance of your findings. There is a need for a more in-depth discussion regarding why thoracoscopic surgery yielded better results than open surgery in your study.

Our results indicate that in our unit, MIS has been introduced safely. OP was in fact associated with a much higher wound infection rate than MIS. Conversely, leak rates in the neck were more common in MIS, however these all settled with conservative management. Discussion line 262 to 268.

Reviewer B

This is a compelling study on the introduction of minimally invasive surgery for esophageal carcinoma in a national university hospital. I particularly appreciate the use of benchmarks to assess the results of this process. However, I have some questions and suggestions for the authors that I believe can lead to improvements in this paper, giving it a stronger impact.

In my opinion, using the abbreviation 'OS' for open esophagectomy can be confusing, as 'OS' is more commonly used for Overall Survival.

Thank you for this feedback. The OS group has been renamed to OP in all the manuscript and tables (highlighted in yellow).

In Table 2, you correctly differentiate the tumors based on location (upper esophagus, middle esophagus, lower esophagus, and the three Siewert types). However, in Table 4, the lower esophagus is not included. If no patients were affected by tumors located in the lower esophagus, it could be clearer if you include it in the table with a 0% value.

Thank you for pointing this out. In practice lower third tumours in Malta exclusively relate to short segment barrets and thus lower third tumours that are in fact not related to the gastroesophageal junction (Siewert type 1 and 2) are rare. The epidemiology of this disease has not been studied in our population but a similar trend has been observed in other southern mediterranean countries [1]. While there have been studies that segregated this group[2] others have group this into distal half[3].

We have clarified this in the paper and acknowledged this limitation (i.e no differentiation between lower third and (Siewert type 1 and 2)). Discussion line 290 to 297

In Table 4, is the stage classified as clinical or pathological?

Table 4 - Stage is pathological. Table 4 has been updated.

After 2017, did all patients undergo MIS? If not, how was the decision made between treating a patient with open surgery or minimally invasive surgery (MIS)?

After 2017, only 2 patients underwent OP (due to COVID restrictions on laparoscopy at the time). This has been added to the Results on page 8 first paragraph, highlighted in yellow line 180 -191.

It's not clear if you performed a matched analysis or not. If yes, it may be necessary to explain it in the statistical analysis paragraph and provide the results of both analyses (before and after matching).

No matched analysis was performed. We unfortunately lack the statistical experience in this technique and feel that the small accrued numbers preclude this too. We have thus removed the term "matched" and acknowledged this limitation in the Discussion line 277-278.

In my opinion, there is no reason to exclude a patient from the analysis even if he dies in the OR.

The available data for this patient has in fact been included (i.e in the 30 and 90-day mortality as well as in the intra-operative hemorrhage data). Hence rather than being excluded, there is no postoperative data. Including this missing value in the postoperative complication data (e.g. for leaks) was felt to dilute these complications when in fact the patient did not have the opportunity to suffer these complications. This would result in better complication outcomes due to a larger denominator. If the reviewers feel it more appropriate to do so, we can include this patient in the denominator for postoperative outcomes.

Is there any difference in the ICU length of stay between the two groups?

This has been tabulated in table 4 and included in the text and abstract. The difference has not been found to be significant on Mann-Whitney U test following parametric testing.

It could be interesting to know if there are any differences in the severity of leaks between the two groups.

Leak severity - in MIS the leaks were all cervical and conservatively managed. In OP 2 thoracic leaks were severe, one leading to death from a stent complication and the other one to sepsis and a protracted hospitalisation. This has now been clarified in 262 to 268.

Reviewer C

Thank you for this audit of 10y of evolving esophagectomy practice in Malta.
- what were the criteria for resection? Even with a catchment of only 500,000, an average of 9 resections per year is very low
With regard to the criteria for resection. Resection was offered to patients fit for prolonged anaesthesia with negative peritoneal cytology, and no evidence of T4b disease or distant metastases on CT, MRI or PET-CT. This has now been clarified in the text (Page 5 in the Staging paragraph highlighted in yellow, line 118).
With regard to the operative numbers. This data was collected from the single national university hospital. Although there is a smaller hospital on the smaller sister island in the archipelago and private institutions, these units are not equipped to carry out oncological oesphagogastric surgery, thus we do not envisage any cases occurring elsewhere. In fact, data, published recently but based on 2012 reports an incidence of n=10 adenocarcinomas and n=10 squamous cell oesophageal cancers that year in Malta[4], which results in similar numbers (considering not all cases are operable at presentation).
The reviewer does however raise an interesting point regarding the epidemiology of the disease and its incidence in a mediterranean country. Assessing the factors related to this relatively low incidence would require a cohort study unlike this here presented

today and may also require consideration of socio-economic factors (recent economic and population growth on the island has been driven by the financial services and software sector, which attracts a younger population). We have elaborated on this point and have acknowledged this limitation (line 290 to 297).

- it is surprising to see no difference in outcomes or survival - one would expect that improvements in practice over time would show this, certainly this has been the case in most other longitudinal studies - can authors specify what types of neoadjuvant therapy were given? The introduction of CROSS and FLOT should have led to improvements in outcome, cPR, R0 rates, and survival. Why do authors think this has not been the case?

With regard to the difference in outcomes or survival, this is in fact an interesting observation. The two cohorts have not been prospectively powered to show these differences and larger numbers (e.g. n=366 for CROSS and n=716 for FLOT) would be needed for use to detect a significant difference. We have now articulated this limitation in the discussion on page 12, line 288 and 289.

- benchmarking is generally a specific term and derived from established methodology. It is unclear to me what the purpose of the benchmarks is in this case and how they have been derived.

This clinical audit seeks to not only disseminate the outcomes of the introduction of the service but also to disseminate the governance of introducing this new operation. In the absence of national set key performance indicators, benchmarks here were derived from landmark studies and international guidelines (Table 3) so as to provide a comparison of our performance. A suboptimal performance when compared to other studies would for example trigger us to halt the service and refer our patients for treatment overseas. We have now clarified this aspect throughout the text including the introduction (line 87 and 88) and discussion (line 225- 228).