Peer Review File

Article Information: https://dx.doi.org/10.21037/jgo-22-633

First round of review comments

Comment1: First, for the title, it is suggested the authors to indicate the comparisons between trimodality and bimodality and the outcomes of interest of this study, as well as the clinical research design of this study, i.e., a retrospective comparative cohort study.

Reply: Please see below for the updated title and running title. A revised title page is included with submitted documents.

Changes in text:

- Title: 'CROSS'-ing into the 'Real World': A Retrospective Cohort Study of Patients Receiving Trimodality and Bimodality Therapy for Esophageal Cancer
- Running Title: 'CROSS'-ing into the 'Real World'

Comment 2: Second, the abstract is not informative and needs some revisions. In the background, the authors did not indicate the clinical needs for the comparisons between trimodality and bimodality and have comment on the strengths of real-world data. The objectives of this study were not described. In the methods, the authors cannot describe statistical analyses only. Please describe the inclusion of subjects, the assessment of clinical characteristics, follow up procedures, measurements of outcomes of interest of this study, and main statistical methods for comparing the two groups. In the results, please report the HR and P values to support "Only treatment modality was associated with overall survival after adjusting for covariates", including the reference group. Because of the small sample size, the conclusion should be made with cautions.

Reply: Please see below for integration of this feedback into the text of the abstract.

Changes in text:

- Background edited to more clearly indicate clinical question presented and objectives of the study. (see Abstract Page 1 lines 1-12). We state we utilized a real-world dataset, word count limitation precludes explicit discussion of benefits of real world data in the abstract, given other requested revisions and existing abstract components.
- Methods revised to discuss inclusion of subjects (Abstract Page 1 line 13-15) and retrospectively recording patient follow-up (Abstract Page 1 lines 17-18). Measurement of outcomes of interest were described in initial draft, along with measurement of clinical variables for association with BMT.
- Results revised to include HR for overall survival cox regression and reference group (Abstract Page 2 line 29). Statement of OS was revised to be 'higher' rather than 'benefited' to avoid causal language (Abstract Page 2 lines 25 ,27).
- Conclusions revised to state observed difference in survival and avoid implying causality (Abstract Page 2 lines 32-33).
- The following components were removed to ensure that the abstract was below 400 words: Comment on radiation dose (Abstract Page 2 lines 30-31), and some minor changes to wording throughout.

Comment 3: Third, in the introduction of the main text, the authors have described the indications of as bimodality "patients who would not tolerate surgery or have inoperable disease," so patients receiving trimodality and bimodality treatments have systemic differences that cannot be adjusted via statistical analysis. The authors should consider whether the proposed comparisons between trimodality and bimodality is appropriate and methodologically acceptable. I do not think so. The authors must be aware of the importance of comparability, which is the pre-requisite for comparisons. A further concern is the clear clinical significance of the comparison.

Reply:

We appreciate this helpful feedback and hope to clarify the aims of our study here and in the changes made to the text, as documented below, and to further clarify the clinical utility of our observations.

With respect to the indications for bimodality therapy, our exclusion criteria aimed to narrow our study population to patients with clinically resectable disease on presentation and excluded patients with staging that would have precluded BMT or TMT with curative intent. Thus, 'systemic differences' between patients receiving TMT and BMT were minimized, as patients who had inoperable disease or clearly were not candidates for therapy with curative intent were not included in our cohort.

In the population studied, as described by these exclusion criteria, most patients receiving BMT planned to receive TMT at the initial timepoint of observation. Following multidisciplinary input, many patients are not clearly optimal surgical candidates nor is surgery contraindicated or excluded – these are the patients that are unlikely to meet criteria for RCTs and for whom this study may be most relevant. For patients who are aiming for surgery but do not ultimately receive TMT due to dynamic and multifactorial decision-making, current literature is focused on primary reasons for nonadherence, and there is a paucity of outcomes data in the literature.

Our analysis provides a multifactorial look at modality and nonadherence and, by identifying factors associated with bimodality treatment, we aim to characterize and describe observed differences in disease, comorbidity, and other clinical factors that may benefit clinical assessment across multiple timepoints during cancer-directed therapy, given that included patients are all presenting with clinically resectable disease and receiving treatment with curative intent. Identified factors observed to be associated with treatment modality require further characterization to better understand how we might support patients to receive optimal therapy as demonstrated through CROSS trial and subsequent analyses. As we learn more about the relative efficacy of 50.4 Gy compared to 41.4 Gy in the setting of bimodal and trimodal therapy for the curative treatment of esophageal cancer in patients, understanding factors associated with treatment modality will be informative when aiming to provide the standard of care for patients for whom surgery is not appropriate or patients who decline surgery.

In this study, we quantitatively describe observed differences in the populations and we agree with the reviewer that causality is not an appropriate conclusion given the methods applied in this study. We have utilized language to avoid statements that explicitly state or imply causality, and any changes made to promote this clarification are outlined in the edits below.

We provide additional context for these observations by providing observations of survival in our cohort, divided by treatment modality, which generates useful data for understanding of possible paths through treatment and survivorship for patients, including those who are in the "grey area" of surgical candidacy at the time of treatment planning. We also provide additional observations of patients declining TMT (but were clinically appropriate for surgery following nCRT), which suggests that longer-term outcomes are similar to other BMT patients, and further study of surveillance following BMT is underway, but may be limited due to inclusion/exclusion criteria of ongoing trials. Ultimately, some

decision-making is patient-driven and we hope to provide data to assist with shared decision-making surrounding treatment modality. Of additional note, there may be patients receiving care at community practices who do not receive early multidisciplinary input, and our data underscores the importance of early multidisciplinary assessment to support optimal therapy, and to assess for surgical candidacy.

In terms of survival outcomes, this study does not aim to provide comparative effectiveness data, as we address below in the next item, and we agree this is not an appropriate aim based on the methodology of our study. We aimed to observe outcomes in our population divided by treatment modality. As mentioned above, there exists little data to characterize outcomes for patients nonadherent to TMT, and we offer additional insight through our analysis.

Changes in text:

- We have clarified that bimodality therapy includes patients declining surgery (page 1, line 5)
- We have indicated that patients for whom candidacy for surgery is unclear at time of treatment initiation, but who are aiming to receive surgery, data from prospective analyses is sparse (Pages 1-2, lines 22-25)
- We have adjusted language to decrease remove implicit statements of causality and limited our use of comparative language (Page 2 lines 30 and 35-36, Page 7 lines 142-147 and 153-159, Page 14 lines 310-311). The title of Table 3 has also been updated with similar changes to language (see table file).
- We have added a statement in limitations to discuss the design of inclusion/exclusion criteria to address limitations of our statistical analysis, as discussed above (Page 13-14, lines 293-296).

Comment 4: Fourth, the methodology of the main text should be organized according to subtitles: patients, assessment of clinical characteristics, interventions, follow up, assessment of efficacy and safety outcomes, and statics. It is also necessary to report the clinical research design and the estimation of sample size, which seems too small for a real world study. For statistics, it is suggested that the authors only describe the characteristics and prognosis of the two groups, not to make comparisons. As commented before, the comparisons between two heterogeneous groups, by nature, cannot generate any meaningful findings, which would be misleading; unless the trimodality and bimodality treatments are interchangeable for the two groups of patients.

Reply:

We appreciate this useful feedback. Our methods have been re-organized as requested. We believe that "statics" was meant to read "statistics," given the subsequent mention of "statistics" in the comment, and we have represented this accordingly in our submission, but have adjusted language to "statistical analysis" due to this inclusion in the checklist provided by the editorial office below.

For logistic regression assessing OS, this analysis was completed in the whole cohort of patients with clinically resectable disease, and modality among other variables for association with survival. As included in our edits above, we have revised language in the text that Kaplan-Meier assessments are not assessing comparative effectiveness, but the observations in our cohort, which includes a significant association between modality and OS, as well as significant associations found on our multivariable logistic regression for association of clinical variables with modality. There is strong RCT data to support superiority of TMT in terms of survival (as compared to surgery alone), and we aim to make observations of survival in our real-world population in the context of these RCT data. We have adjusted our language

to avoid stating that we are "comparing" our observations, and we are rather observing the outcomes in these two sub-groups of our population.

In terms of the size of our study, we were appropriately powered to detect approximately a 30% difference in 3-year survival (calculated based on r = 0.05; beta = .2 and survival of approximately 60% vs. 30%) which would require 40 per group and a sample size of 80. Given that the actual difference in survival in a 3-year time frame between comparable patients receiving trimodality and bimodality therapies has not been established, we assessed the data available for all patients meeting our exclusion criteria and did observe a significant effect on overall survival. Our analysis was likely underpowered to detect factors with a smaller magnitude impact on survival on multivariable regression, and this is an indication that further utilization of real-world data stands to benefit our understanding of esophageal cancer therapy.

Changes in text:

- Methods and materials have been re-organized as requested using the suggested headings, please see document with track changes for full revisions (pages 2-5)
- Additional text was added to clarify and complete sections of 'assessment of efficacy and safety outcomes' and 'follow-up' (Page 4, lines 70-78)
- Limitations edited to include statement on the power of our analysis (Page 14 lines 300-302)
- STROBE checklist updated to align with sub-headings in materials and methods.

Re-review comments

Reviewer A

The authors present their experience with management of non-cervical esophageal cancer. Is a 10-year experience that includes almost 100 patients. I have multiple comments.

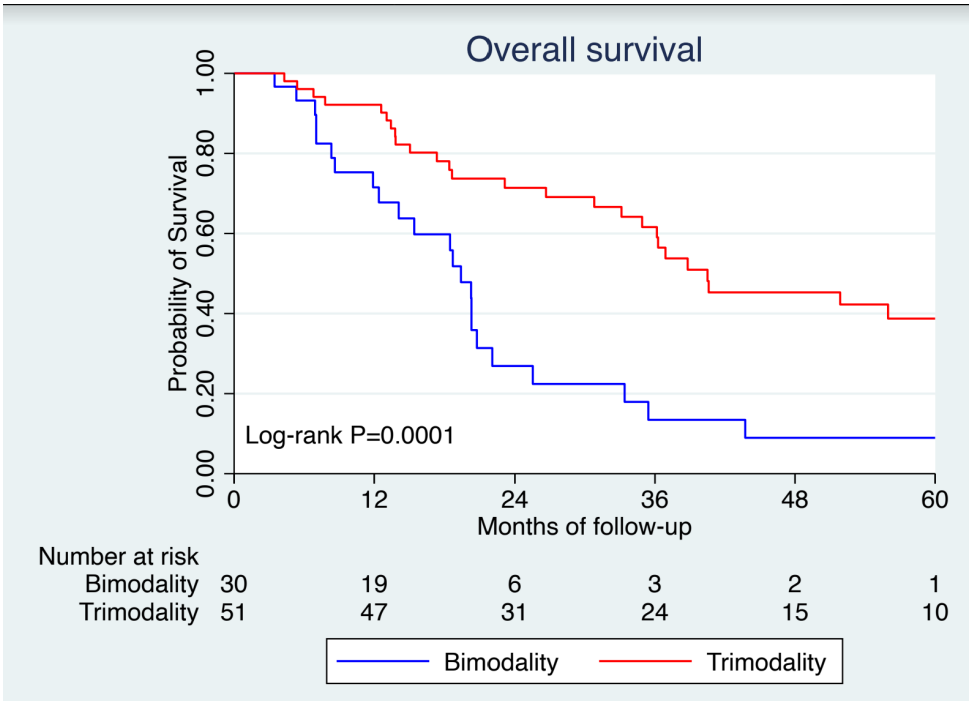
1. The tittle needs to be changing this is not a comparison of a cohort of patients treated under Cross protocol that did or did not have surgery this is a descriptive study evaluating outcomes of patients receiving chemoradiation followed by surgery versus patients receiving definitive chemoradiation.

- Reply: We hoped through the original title to draw attention to the aim of our analysis to include patients who are typically excluded from RCTs such as the CROSS trial, which is currently one of (if not the singular) most important trials determining our standard of care for esophageal cancer. Of three reviewers, we have received one comment about the current title. While we would prefer to keep the current title because we feel it will best catch the attention of potential readers so they may engage with our real-world data analyses, we offer the two alternative titles, which can be utilized at the discretion of the journal and reviewers and without objection from the authors in place of the current title:
    - Who Reaches Esophagectomy in the 'Real World'? An Observational Retrospective Cohort Study of Patients Receiving Trimodality and Bimodality Therapy for Esophageal Cancer

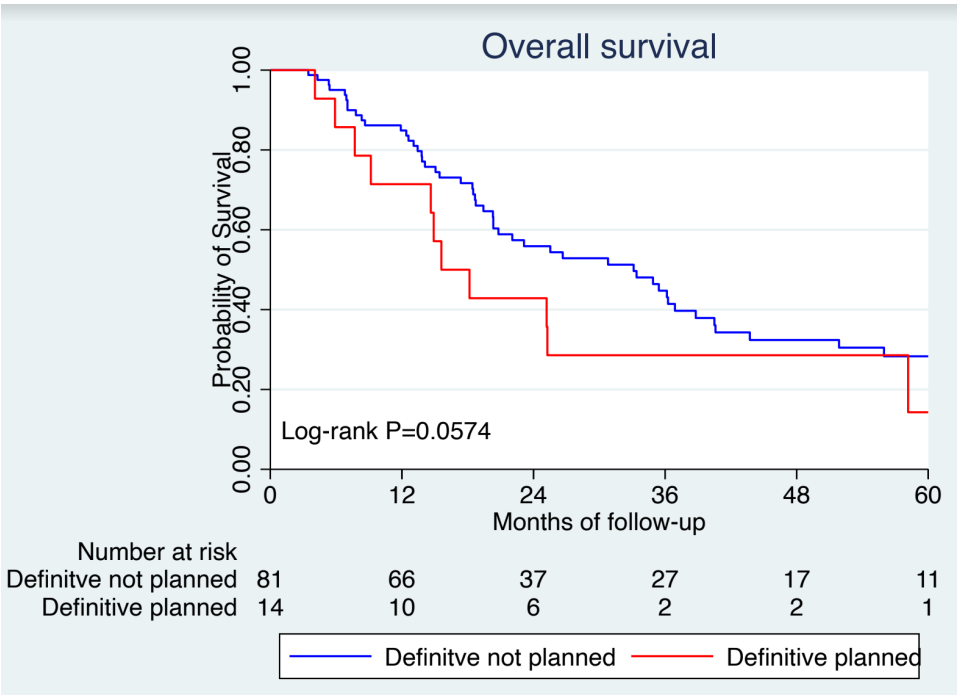
    - An Observational Retrospective Cohort Study of Consecutive Patients Receiving Trimodality and Bimodality Therapy for Esophageal Cancer
- Changes in text: See proposed alternative titles above.

2. Survival analysis should start based on intention to treat, or the original consideration, neoadjuvant versus definitive therapy and with that starting point we can be closer to fully understand the different groups. Once that is defined, we can see how many people crossed over and what was their outcomes as is different to offer surgery after favorable response to therapy

- Reply:
    - With regard to crossover, we have examined this at length in our section on adherence to trimodality therapy. For these analyses, we recorded a variable that allowed us to exclude patients for whom definitive bimodality therapy was documented as the plan from the beginning of therapy, and thus this analysis does follow the intention of treatment as documented in the patient's medical record. We did not record the specific reason definitive BMT was planned for each patient. There was some variation between patients, especially those being referred to our center from our catchment area, with regard to what point in time they were seen by a surgeon to discuss esophagectomy. For this reason, although we have data representing the documented intent to deliver definitive BMT at the beginning of treatment, the variation in time after diagnosis and time relative to therapy when patients saw a surgeon does introduce variability that was not accounted for with this variable. For many patients, contingencies were documented rather than explicit plans for definitive therapy, which we were able to characterize in our analysis of adherence. Hypothetically, prospective analysis of intention to treat in a larger cohort is better suited to assess any difference in survival between comparable patients based on intention to treat.
    - To further assess the intention to treat with definitive BMT in our cohort, we have utilized the previously discussed variable to create additional Kaplan-Meier analyses of OS with the exclusion of the 14 patients for whom definitive BMT was planned from the beginning of therapy, with results similar to our original analyses.

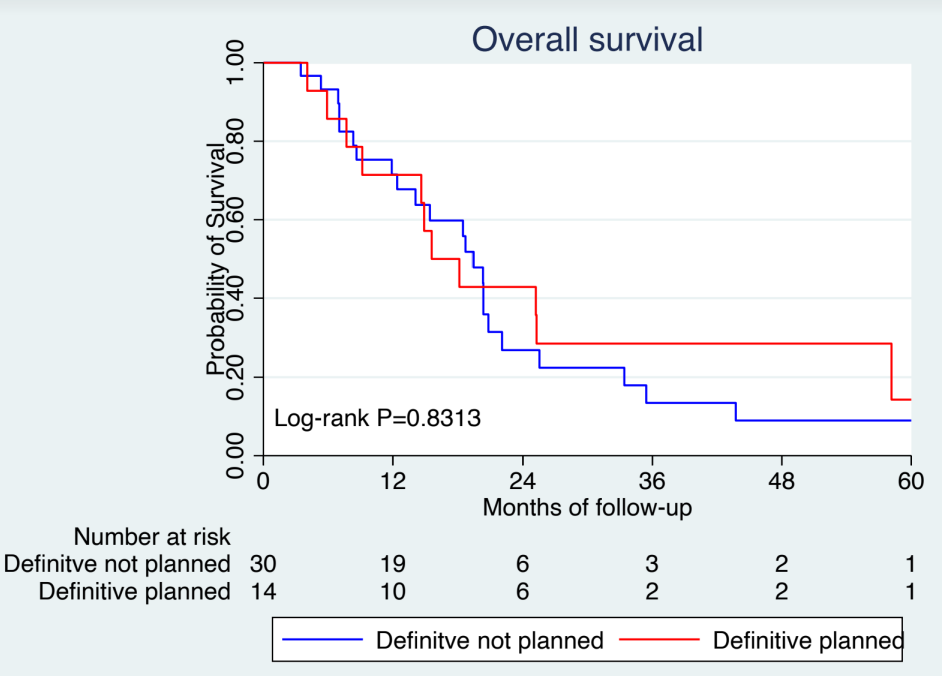


- We have also completed a KM analysis based on our recorded variable comparing patients with documented plan for definitive BMT to all other patients. This demonstrated a difference in OS approaching significance, which we feel is expected for a comparison of BMT patients to a cohort including TMT and BMT patients.



- When we assess documented intention to treat with definitive BMT only among patients who received BMT, there is no significant difference in overall survival. Given that these are patients with comparable outcomes in terms of survival, we believe that our selection criteria have identified a population for whom consideration of TMT is appropriate, and this is functional as an original

consideration with regard to modality in the scope of retrospective analysis. Nonadherence to TMT and/or intent to treat with definitive BMT occurs for a variety of reasons, and based on the information below, we believe we have evidence that the data we do have available on intention to treat with definitive BMT does not contribute an additional or clinically useful element to our survival analysis given the consideration of our inclusion/exclusion criteria.



- We would like to underscore that our exclusion criteria were designed to capture patients with clinically resectable disease at initial assessment, and we believe that the design of our exclusion criteria does address a plurality of the considerations that affect the original consideration of neoadjuvant therapy with the exclusion of factors that this study was specifically intended to characterize, including age, comorbidity, and performance status. In a multivariable logistic regression of these factors for association with modality, excluding the 14 patients for whom definitive BMT was planned delivers similar results to our original analysis, as shown below.

Logistic regression — Number of obs = 60; LR chi2(8) = 26.82; Prob > chi2 = 0.0008; Log likelihood = -27.339269; Pseudo R2 = 0.3291

modality	Odds ratio	Std. err.	z	P>|z|	[95% conf.	interval]
charlsoncat						
5-7	.1355194	.1160669	-2.33	0.020	.0252919	.7261414
8-10	.0701604	.0917643	-2.03	0.042	.0054049	.910747
pspre						
ECOG 1	.2614622	.2070434	-1.69	0.090	.0553822	1.234376
ECOG 2	.3141879	.3853601	-0.94	0.345	.0283901	3.477054
nstage						
N1	.9646706	.722443	-0.05	0.962	.2222865	4.186441
N2-N3	.0216161	.0355878	-2.33	0.020	.0008578	.5447014
presentingsxs2						
Other	.1910314	.268335	-1.18	0.239	.0121745	2.997484
chemoheld						
Yes	.088811	.070742	-3.04	0.002	.01864	.4231431
_cons	74.86735	91.16629	3.54	0.000	6.883046	814.3371

  - Based on these findings, we do not feel that it is appropriate to alter our survival analyses based on the variable that was recorded in our study, as it is challenging to discreetly and completely define intent in retrospective real world data. We feel that our exclusion criteria address appropriateness of consideration of neoadjuvant approach, and our previously presented results are in line with analyses utilizing data on intention to treat as it is available in our dataset, as discussed above.
  - We would be happy to include any of the above figures in the supplemental appendix if requested or to incorporate a curve of survival based on our available data on ITT with definitive BMT as one of the components of our figure characterizing OS should this be requested by the journal/reviewers.
- Changes in text: We have added a paragraph to limitations summarizing the discussion above and stating that we feel it is of limited utility to assess survival in our cohort based on intention to treat compared to our initial assessment. As stated above, inclusion of these findings in our supplemental appendix may be appropriate if requested by the reviewer and/or journal.

3. The non-Cross criteria need to be better defined as if we are talking patient not eligible to a trial fulfilling anatomic criteria but having poor PFS 2 is different than non-surgical candidates but T1N0 tumors.

- Reply: The individual published criteria for exclusion in the CROSS trial were considered in our multivariable logistic regression for association with modality (table 3) and our multivariable cox regression analysis for association with survival (see supplemental table 3). Univariable analysis included CROSS eligibility for each, but this was excluded from multivariable analysis in favor of individual exclusion factors, which we agree provide more precise information. A Kaplan-Meier curve is included that presents survival of only CROSS-ineligible patients based on modality, while this

does composite the individual criteria, we believe demonstrating that the CROSS inclusion criteria do not fully account for individuals who would benefit from TMT is a reasonable and clinically important point that can be explored through our dataset.

- Changes in text: Discussion, 'survival benefits of TMT' paragraph one now includes a sentence to explicitly recognize composite criteria utilized in cox regression of TMT and BMT in CROSS ineligible patients, and its utility, as above.

3. All the results seem to indicate that any efforts to not operate on those patients are associated with dismal survival. All commentary about patient preference for non-surgical approaches need to be contextualized on the fact that the patients declining surgery had a much worse survival than surgical patient and almost same as patient that were not candidates for surgery.

- Reply: We agree that survival is an important context for patient preference, especially for patients for whom survival is the primary goal of care. This is a complex decision, involving the weight of permanent and temporary morbidities. Further sociocultural factors influence patient's outlook on surgical management and can further complicate shared decision-making. Our survival curve comparing patients who decline surgery and TMT patients, to which I believe the reviewer is referencing, provides important context in our discussion. In the discussion section "drivers of TMT nonadherence" we discuss literature that demonstrates patients have concerns over quality of life and preference for nonsurgical management up a certain reduction in mortality benefit. We also discuss the timeline of recurrence in studies of surveillance after BMT, and the impact on ability to complete esophagectomy. We have made changes to this section, as below, to better emphasize these points.
- Changes in text: in the discussion section "Drivers of TMT Nonadherence," we have added explicit comparisons between our data and available information from the literature about acceptable differences in survival for patients to elect non-operative management. We have added a paragraph break for the structure of this section to better separate points regarding (1) prior observations of nonadherence due to patient choice (2) magnitude of acceptable impact on survival compared to observed data from a patient perspective and (3) available data on active surveillance after dCRT, including context of histology-specific data.

4. The paragraph about Low radiation doses is not appropriate as this was not what was delivered to the patients in the series as seems to be just a philosophical point, not based on their actual data.

- Reply: The practice at our institution had transitioned over the study period to treat primarily with 4140 as utilized in the CROSS trial. Direct comparison of 4140 and 5040 is lacking, but also an area that will need to be explored in the context of which patients are most likely to receive or benefit from TMT in addition to any RCTs that may be completed. In this section, we hoped to both reflect our changing practice and to discuss future directions related to dose that may help to mitigate treatment toxicities we observed to be associated with treatment modality and to facilitate surgery in appropriate patients. Due to the sample captured in our study period, we were unable to comment more on our presented data, but we believe this is an important future direction from our study.
- Changes in text: We have shortened this discussion to include the most salient points,

as above. Please see the changes in the text.

5. Statistical analysis needs to be revised as some inconsistent data is provided such a very long follow up period with patients with dismal survival. Concerned about statistical analysis as they are all over the place, using different denominators and different criteria to define the cohorts. The median follow up was 69m with a median survival of 26m that seem to be a mistake.

- Reply:
    - The median follow up was greater than the median survival due to the employment of the reverse KM method. Please see our citation for this method with more information on how these analyses are censored to compute these results.
    - We appreciate the detailed review of our findings; however, without further precise comment on individual aspects of our analyses, I will discuss the anticipated areas of our analysis that I believe the reviewer is indicating.
        - Generally, the difference in denominators for various datapoints reflects the availability of data in this real-world retrospective study.
        - We included patients who did not plan for TMT but met all inclusion criteria, thus the non-adherence analysis excluded patients for whom nonadherence was not applicable.
        - For survival analyses, we did include analyses of subsets of the cohort, as in the comment above regarding CROSS-ineligibility. Additionally, we included analysis of patients who were offered surgery after completion of nCRT but opted for non-operative management, as this was a population of interest identified in our study and a population for whom real world data needs to be further assessed to inform shared decision-making between physician and patient.
- Changes in text: None

Reviewer B

This is a well-written manuscript which aims to answer an important real-world question. While patients with locally advanced esophageal cancer are recommended to undergo tri-modality therapy, in practice these patients commonly carry comorbidities which exclude them from surgical candidacy. It is valuable to better understand the outcomes of patients who ultimately undergo BMT compared to TMT. Again, the manuscript is very well-written and the statistical analysis is sound.

- Reply: Thank you for your positive feedback on our statistical analysis and the writing contained in our manuscript. We agree that modality is importantly addressed as a 'real-world' question in supplement to data from RCTs, to inform treatment of persons for whom candidacy or desire for operative management varies.
- Changes in text: None

I believe the citation for the reverse Kaplan Meier method is missing

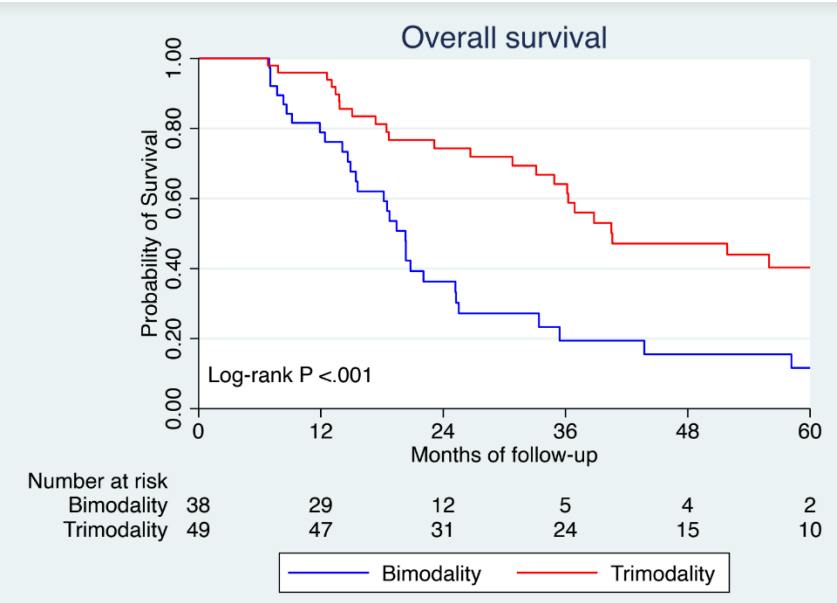
- Reply: Citation 15, [Clark TG, Bradburn MJ, Love SB, et al. Survival Analysis Part I:

Basic concepts and first analyses. Br J Cancer. 2003;89(2):232–8.] was included to cite the reverse KM method. Although not the sentinel description of the method, it discusses methods and utility of this approach. We have added a more sentinel analysis for this method as a second citation.

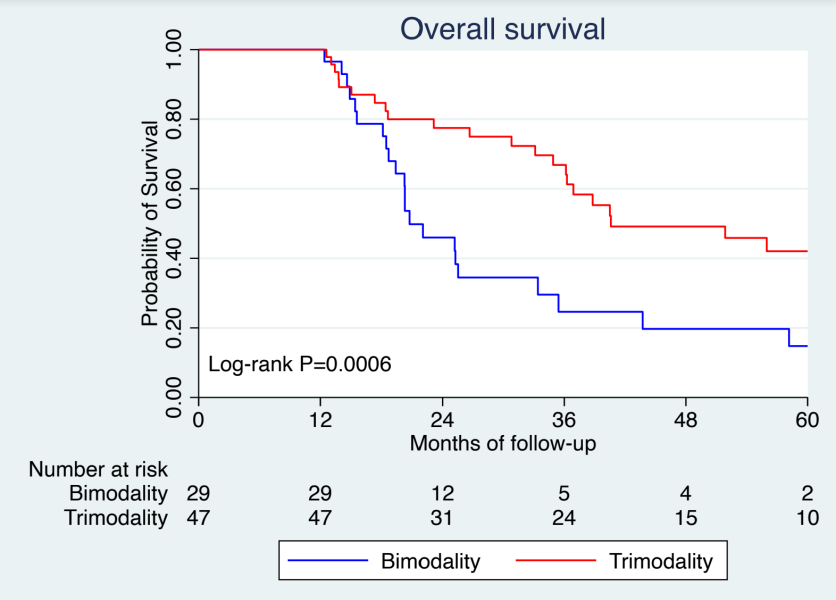
- Changes in text: citation of Schemper et al.

Were there any patients in the BMT group who were intended to undergo surgery but died during preoperative therapy? If so, this introduces immortal time bias that favors the surgery arm. It may be difficult to fully determine this from a retrospective review so this should be acknowledged in the limitations.

- Reply: This is an important detail that is helpful for us to include; thank you for including this in your comments. Two patients passed during anticipated duration of radiation. One of two planned for definitive radiation at the start of treatment. One patient died approximately 3 weeks after treatment and planned for TMT. The latter two patients account for patients for whom "expired" is the documented reason for nonadherence. This bias should be limited to association of modality with survival. We completed 6 and 12 month landmarked analyses as sensitivity analyses, which are included below. We continue to demonstrate similar findings in these analyses and we believe these demonstrate minimized potential effect of immortal time bias in our study.



(Above is 6-month landmarked analysis)

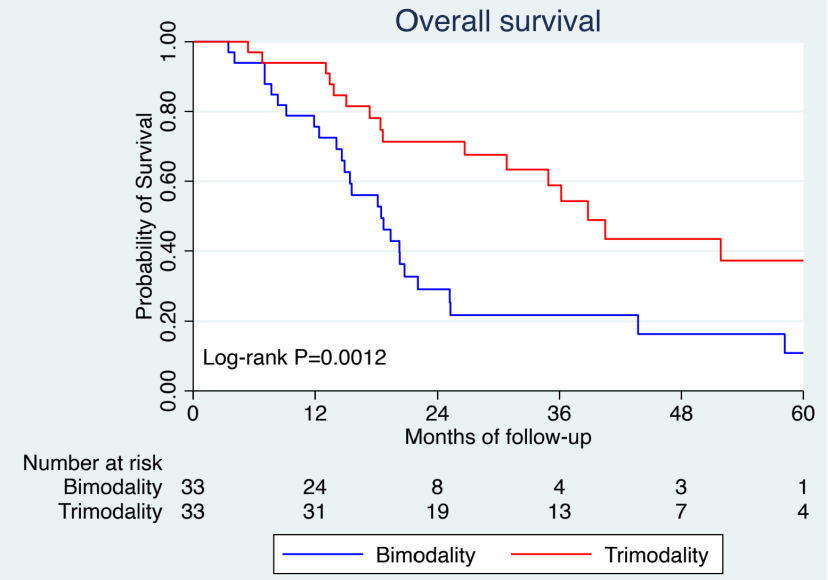


(Above is 12-month landmarked analysis)

- Changes in text: We discuss these sensitivity analyses in our results, section 'survival' second paragraph. Brief paragraph added to end of limitations to address immortal time bias. We have agreed to open publication of peer review file, so these analyses should be included herein, but we are happy to format for addition to the supplemental appendix if desired by the journal and reviewers.

The use of induction chemotherapy prior to chemoradiation also introduces a time-varying factor. It is understood that there was no statistical difference in the rate of induction chemotherapy between BMT and TMT arms. Still, can the authors provide a sensitivity analysis excluding these patients?

- Reply: Thanks for your feedback on this point. Firstly, our definition of time included day #0 as the date of diagnostic biopsy. The time variance for patients is multifold and complex with regards to date of starting therapy and date of surgery compared to biopsy. In addition to the evidence that distribution of patients between arms was not significant, we also have assessed OS based on induction and found no difference (log rank P=0.459). We have completed a sensitivity analysis with a Kaplan-Meier curve that excludes patients with induction chemotherapy with similar observations to the whole cohort as requested, please see below.



- Changes in text: None

Ultimately, the patients in the BMT arm are not surprisingly significantly older with worse performance status, and to some degree it is difficult to completely make up for this with multivariable adjustment as the patients are inherently different. This is actually the goal of the study - to better understand real world outcomes- but still limits conclusions. This should be addressed in the limitations and conclusions.

- Reply: We agree that there are underlying differences between the cohorts based on treatment modality that may be hard to adjust for in a real-world retrospective study. We agree that our description of real-world outcomes was a central part of our approach, and we appreciate that this came across clearly in our manuscript. We believe your point is captured in our limitations section: *"Through multivariable analysis, we controlled for differences in common prognostic factors between patients receiving TMT and BMT when analyzing for association with survival and other outcomes. Additionally, our exclusion criteria were designed to create a cohort in which patient presentation was both inclusive of patients who were older and with increased frailty, but minimized uncontrollable variation in disease through exclusion of patients receiving surgery without curative intent. However, some patients may have experienced inferior outcomes due to factors for which we did not collect data or had incomplete data, or had complexities difficult to characterize and study in a single-institution retrospective cohort."*
- Changes in text: We have further clarified limitations through edits of the section above to ensure that these limitations are discussed explicitly in relation to the conclusions made through multivariable analysis. Please see limitations, first paragraph, final sentence.

Reviewer C

Below are some general thoughts about the article. There are further comments in the PDF.

- This is a good retrospective study of real world data.
    - Reply: Thank you, we appreciate this positive feedback.
    - Changes in text: none
- It is very well written
    - Reply: Thank you, we appreciate this positive feedback.
    - Changes in text: none
- This retrospective cohort has a very significant proportion of patients who did not proceed to surgery. From my understanding, there is no intention-to-treat analysis resulting in overly significant superiority of TMT vs BMT in OS. (Apologies if missed this in the statistical analysis.) There are bigger datasets which show non-inferiority in some circumstances such as patients with SCC of the esophagus. The BMT cohort also has a lower than expected OS and this is not clear why. This study fails to discuss the use of peri-operative chemotherapy that is commonly used in GEJ tumours.
    - Reply: In the comments below we precisely address comments on intention to treat analyses, overall survival observations in the BMT cohort, and the use of peri-operative chemotherapy, as represented by comments on the PDF of the article. For some of these discussions, we will refer to comments addressed from reviewers A & B.
    - Changes in text: Please see further discussion below.
- As a result of the points above, there is bias in the reported data and much of the discussion/conclusion section can be challenged.
    - Reply: We recognize that our retrospective single-institution analysis has limitations due to the design of study, which we have addressed in our limitations section, and as edited based on all reviewers' feedback. As stated above, please see our response to the individual points based on the comments that were provided on the PDF. Although limitations exist and are noted, we believe we provide additional and clinically relevant context through reporting real-world outcomes in a wide scope of patients.
    - Changes in text: As otherwise noted based on individual comments.
- The discussion needs to look ahead at a changing paradigm, there is the factor of post-op immunotherapy to consider, and also other technologies such as proton beam treatment which may shift goalposts again.
    - Reply: We address two changing paradigms in our discussion that we feel reflect the current trajectory of treatment for non-cervical esophageal cancer. Firstly, we discuss the changing recommendations surrounding dose in our section discussing LDRT, although this has been trimmed based on the feedback of Reviewer A. We also discuss adjuvant nivolumab, which was introduced in our practice after the study period based on the CheckMate577 trial and reference ECOG 2174 which is evaluating neoadjuvant nivolumab. We appreciate the feedback for inclusion of proton beam treatment and have included additional discussion as below. It is notable that proton beam treatment may be of limited accessibility to individuals living in rural areas and/or individuals of fewer resources, and in our real-world cohort it would likely be less accessible than addition of systemic agents, as above, or lower dose radiation to decrease toxicity based on current availability by geography.
    - Changes in text: Please see additional discussion of proton beam therapy in discussion

'clinical variables associated with BMT' final paragraph with citation to Lim et al. findings of decreased toxicity vs. IMRT.

I would suggest reviewing the comments in the PDF. There will be some major revision required to the discussion and tweaks elsewhere prior to this being acceptable for publication. I would suggest running an intention to treat analysis. I don't think the lesson from this data is so much that TMT is so far superior to BMT. I don't think this data answers this question. As a side note, CROSS showed superiority of TMT to surgery alone (not BMT), It is rather that patient selection and patient buy in is vital. In appropriate patients, there is little doubt that TMT is the gold standard treatment. I think the challenge is to define that patient group.

- Reply: We agree that identifying patients who can tolerate TMT and benefit from it is imperative. We also agree that our knowledge of TMT as a standard treatment due to survival benefit is not our primary lesson from the data. Our goal was to examine a real-world population to determine how many patients completed TMT and if there were factors associated with TMT completion in clinically resectable patients. We then characterize survival in the group examined for the former questions to further characterize the study population and to compare to data collected in different, generally healthier, populations. We discuss your comments on the PDF each individually as noted below.
- Changes in text: See below.

Replies to comments on PDF:

1) line 2: A standard of care rather than the standard of care. Perioperative chemotherapy (e.g. with FLOT) is also considered a standard of care in in gastro-esophageal cancers).
    a. Reply: Please see comments below for further discussion of perioperative chemotherapy. We recognize the use of perioperative chemotherapy in specific contexts discussed above and have made the requested changes, as below.
    b. Changes in text: Replaced "The" with "A." We have also modified this same language in the abstract.
2) Line 12: This is true. However, are trials comparing NA CRT vs Perioperative chemo showing non-inferiority. Admittedly these studies include cohorts that include gastric cancers but also some with GEJ tumours. e.g. NEO-AEGIS Neo-AEGIS (Neoadjuvant trial in Adenocarcinoma of the Esophagus and Esophago-Gastric Junction International Study): Preliminary results of phase III RCT of CROSS versus perioperative chemotherapy (Modified MAGIC or FLOT protocol). (NCT01726452). John V. Reynolds, Shaun R. Preston, Brian O'Neill, Maeve Aine Lowery, Lene Baeksgaard, Thomas Crosby, Moya Cunningham, Sinead Cuffe, Gareth Owen Griffiths, Rajarshi Roy, Stephen Falk, George Hanna, Frederick R. Bartlett, Imelda Parker, Alberto Alvarez-Iglesias, Magnus Nilsson, Guillaume Piessen, Signe Risum, Narayanasamy Ravi, and Raymond S. McDermott Journal of Clinical Oncology 2021 39:15_suppl, 4004-4004
    a. Reply: We recognize the use of peri-operative chemotherapy in GEJ tumors, and it is our institutional practice to offer peri-operative chemotherapy for management of Siewert III tumors, which were excluded from our analysis (see below for changes to methods to clarify this point). Our institutional

practice is for Siewert I-II patients to receive neoadjuvant chemoradiotherapy followed by esophagectomy, based on the category I preferred recommendation of the NCCN for GEJ tumors. Per the NCCN, "preoperative chemoradiation is the preferred approach for localized adenocarcinoma of the thoracic esophagus or EGJ. Perioperative chemotherapy is an alternative option for distal esophagus and EGJ." Neo-AEGIS results do demonstrate non-inferiority for overall survival following peri-operative chemotherapy, as in the abstract referenced and recently presented at the 2023 ASCO GI symposium. Of consideration, "In the MAGIC/FLOT arm, 82% of patients achieved R0/negative margins compared with 95% in the CROSS arm." The full clinical significance, including stratified significance by Siewert classification is not yet known. Based on published material assessed by our team, we were unable to find published distribution of tumor location in the study, which may limit applicability to our study population which excluded Siewert III. Furthermore, the protocol does not appear to provide adjuvant immunotherapy for patients receiving TMT, and comparison of perioperative chemotherapy and TMT with adjuvant nivolumab needs to be completed before concluding noninferiority of current standard of practice, which is pre-dated by our study and the NEO-AEGIS protocol.

   b. Changes in text: We include a statement, reflecting comments above, on the NEO-AEGIS protocol initial findings in Discussion 'Survival benefits of TMT' final paragraph.

3) Line 55: Ideally this should be classified according to AJCC GEJ type. Typically Type 1 and 2 GEJ are classified esophageal tumours while Type 3 treated as gastric.
   a. Reply: We agree that added precision is helpful for defining these criteria. We defined our exclusion criteria based on Siewert type, for the reasons you mention (and per our discussion of variation in treatment above) and will modify our text to include this explicitly.
   b. Changes in text: We have explicitly defined our exclusion criteria based on Siewert type in materials and methods section entitled 'patients' paragraph two. We have also added a citation for this definition.
4) Line 124: 37% dropout rate is very significant
   a. Reply: We agree. Dropout rate was 22.8% when excluding patients who were offered surgery but declined, which is also clinically significant. As mentioned elsewhere in comments, we found the proportion of patients declining surgery to be an interesting and unexpected finding. In our discussion of nonadherence, we believe we flesh out our thoughts on this finding and the overall dropout rate, and have addressed additional comments as below. Please also see Reviewer A comment #3.
   b. Changes in text: None
5) Line 139: This feels quite low. RTOG 8501 had higher 5 year survival (27%). The UK SCOPE1 has nearly 60% 2 year survival
   a. Reply: Adenocarcinoma histology was observed in 82% of BMT patients in our study and 28% of BMT patients had an ECOG PS of 2. RTOG 8501 did

not exclude patients with low performance status (minimum KPS 50) but did exclude patients with history of malignancy (we did not and included history of malignancy in our survival model, as below). We don't believe our populations are comparable, both study arms are >85% SCC histology. Upon reviewing the data in the study in greater detail, there appear to be few patients of performance status <70 (accounting for only 8% of randomized BMT patients and 0% of nonrandomized), whereas our cohort of BMT patients was observed to have 28% ECOG 2, which is equivalent to KPS 50-60. The SCOPE-1 trial only included patients with WHO PS 0-1 (50.8% of study population were WHO 0 compared to 25% of BMT patients in our study) and was similarly SCC predominant (72.9%). SCC patients have been observed to have noninferior survival with dCRT compared to TMT, and thus this population is similarly difficult to compare directly to our cohort. As we state in our discussion, Radiation Therapy Oncology Group (RTOG) 8501 demonstrated five-year OS of 13% and only 1 of 23 patients alive at longer term follow-up for definitive BMT in patients with adenocarcinoma histology, which may be more comparable to our population due to distribution of histology in the previously discussed studies.

   b. Changes in text: None

6) Line 145: I can't see how this is the case. BMT use is associated with poorer PS, older cohort, fewer chemo cycles and yet treatment modality is only factor assoc with improved OS. I'm not a statistician but does not seem to reflect the data. Have you looked at baseline PS or number of chemotherapy cycles. Also have you looked at radiation dose as variable affecting OS? While i don't doubt that your stats figures are legitimate, i think the way the data has been analysed, has resulted in an outcome that is not reflective or your data. I think there needs to be intention to treat analysis before you can make this argument, suggest also running additional variables to look at OS.

   a. Reply: Supplemental table 3 contains a more complete picture of our survival analysis. On multivariable analysis, we did adjust for an age-adjusted comorbidity index and history of neoplasm, thus we have adjusted for age. The number of chemotherapy cycles held and performance status, while associated with modality, were not independently associated with survival when accounting for modality and were thus not included in our model. Number of chemotherapy cycles is further complicated in our cohort due to the inclusion of patients receiving induction chemotherapy on a research protocol, as we discuss in the paper. The mode radiation dose was 50.4 Gy in our cohort. Given the single institution cohort, and tendency to treat with the same dose, our study's ability to detect differences in dose and survival was decreased, we suspect a higher powered large-cohort multi-institution study would be necessary to do this retrospectively (or even prospectively). We have examined the mentioned variables for association with survival, and with the inclusion of modality in our cohort, we do not find significant associations with survival. Please see our discussion of intention to treat in response to comments from Reviewer A.

b. Changes in text: None

7) Line 154: This again in quite low, indicating either excellent surgeons, post op support and/or overly cautious patient selection to proceed to surgery. Given that data for survival outcomes for BMT is lower than expected, i think there is room to be more aggressive with regards to surgery post CRT.
    a. Reply: Horne et al (citation 18) reported multi-institutional data with observation of 30- and 90-day mortality was 2.4% and 4.5%, respectively. We observed 30 and 90 day mortality of 2 and 4%, respectively, which is similar to previously observed values, and we believe our findings to be well within the expected variation from values observed in a larger population study.
    b. Changes in text: We have edited this paragraph to state the percentage of patients observed rather than the absolute number to facilitate comparisons to previously observed values.
8) Line 157: I don't agree with the argument in this paragraph. While there is statistical significance the TMT is better than BMT in OS, there is a huge drop out rate from CRT proceeding to surgery (nearly 40%) and therefored there is inherent bias in the statistical analysis. I think there needs to be an Intention to treat analysis based on the initial planned treatment strategy, if you are to make this argument.
    a. Reply: Please see our discussion of intention to treat in replies to Reviewer A, and the stated changes to the text that are included in that reply. Please also see our response to your overall comment above about the primary function of the survival analyses to further contextualize the analysis of modality and adherence in a real world population. The crossover of 37% is well characterized in our analysis of adherence (see reply to comment 12, below regarding adherence) and this analysis excluded patients for whom definitive BMT was clearly documented to be planned at the start of treatment. As included in the reply to Reviewer A above, we do observe a difference in overall survival approaching significance between patients with planned definitive BMT and other patients, but this is an expected finding given the composition of groups being compared. To further address the possibility of immortal time bias for patients for whom tolerance of therapy was quite poor and/or expired during therapy, please see our reply to reviewer B and sensitivity analysis with 6 and 12 month landmark analyses of OS, which we believe also addresses aspects of the concern for bias raised in this comment.
    b. Changes in text: As above in reply to reviewers A and B.
9) Line 186: This is interesting but we know from much larger studies, number of chemotherapy sessions affects the OS.
    a. Reply: Thanks for bringing up this additional element for discussion of our result for our variable recording held chemotherapy cycles. For this assessment, we were specifically evaluating held chemotherapy sessions for association with modality. Held chemotherapy sessions, in a larger cohort, might be reasonably expected to affect overall survival independently from modality, but this was not observed in our cohort. With the addition of adjuvant nivolumab to the current standard of care with OS benefit, the independent

association of the number of chemotherapy sessions with OS will need to be further characterized.

b. Changes in text: See discussion, “clinical variables associated with BMT” paragraph 3 for additional sentence summarizing comments above.

10) Line 192: While i don't agree with anything mentioned in this paragraph, i would argue that there is nothing in the data presented that has relevance to prehab and dietetics input e.g. difference in post-op toxicities stratified to prehab vs no prehab or percentage of patients receiving this. Prehan and nutritional support should be taken as a standard of care.

a. Reply: We agree that prehabilitation and nutritional support should be the standard of care, and we recognize that this is extrapolated as an area of future refinement based on results indicating that many factors associated with BMT are observed on presentation, and mitigation of their effects may require tailored interventions. We have edited this section to the most salient points.

b. Changes in text: we have edited this section and have emphasized that standard of care should include these interventions. We have included language specifying that further exploration of the most effective interventions and patients who stand to benefit is needed, along with any impact on adherence to treatment.

11) Line 241: I agree, and reflective of our practice in the UK for patients in the Neadjuvant setting. I think you can expand on the dose used in the definitive setting (BMT). Our practice is to use 50Gy/25# in keeping with RTOG 8501 and UK SCOPE trials. The question about dose escalation in oesophageal cancer is not yet completely answered despite a few negative trials (INT 0123, ARTDECO). The UK SCOPE2 trial is currently recruiting and testing given 60Gy to a smaller SIB volume. 50.4gy/28# , particularly if given without chemotherapy is much to cool a dose to be 'curative' or even provide durable local control. Little mention is given about rt quality assurance, technique or modality.

a. Reply: Firstly, we would like to clarify that all discussion of radiation dose is predicated on the concurrent administration of chemotherapy, with our study focusing on bimodality and trimodality treatment representing nCRT +/- surgery. We specify that our preference is for 41.4 Gy in patients committed to going to surgery, and have added discussion of definitive dosing as requested, see below. The NCCN guidelines specify 50-50.4 Gy as the recommended dose in the definitive setting and 41.4-50.4 Gy in the neoadjuvant setting. We understand that trials of dose escalation are negative to date and believe this to be outside the scope of our paper, especially when considering the genetic profiles that may be necessary to identify responders to dose escalation. We hope to satisfy both reviewers commenting on this paragraph by covering the most salient points, while also thoroughly addressing each. In terms of quality assurance and modality of radiation, the single-institution design of this study did mean that the vast majority of patients were treated on similar machines with similar quality assurance practices. Most patients received photon-based external beam radiation with 3D conformal planning. Further discussion of

quality assurance and planning, for which we did not collect specific data, is outside the scope of our study, but reasonably considered when discussing treatment toxicity. See additional discussion of proton therapy elsewhere.

b. Changes in text: See discussion section 'clinical variables associated with BMT' second to final (previously final) paragraph. We now discuss dose for definitive treatment.

12) Line 244: this is probably the most interesting part of this study. The non-adherence rate to surgery is very high in this study. It would be nice to get some more detail as to why. Surely some of this must come from clinician steer. Is insurance a factor?

a. Reply: We describe the drivers of nonadherence in detail in results section 'TMT nonadherence.' ["Reasons for patient nonadherence to TMT are recorded in Table 4 alongside findings from two similar analyses of nonadherence (11,12). Among the 12 patients who declined surgery, 6 (50%) patients cited preference for non-surgical management, 3 (25%) patients cited concern for morbidity, 2 (17%) patients cited advanced age, and 1 (8%) patient cited optimism based on response to chemoradiation. One patient with preference for non-operative management also cited the lack of a support system to enable post-operative recovery."]. We captured insurance status at time of diagnosis in our dataset, but did not observe this to associate significantly with modality or survival (we captured public vs. private insurance and lack of insurance). Of the 95 patients included in the study, only 1 patient was uninsured and 1 patient did not have data for insurance status. Contemporary to part of the time period of our study, RTOG 0436 had demonstrated favorable survival outcomes following definitive BMT, which may have influenced the counseling received by patients. We further discuss the rate of nonadherence in comment #4 above.

b. Changes in text: None

13) Line 282: You have to mention potential for bias in the data as no intention to treat analysis. If you selectively pick out the patients who proceeded to surgery, of course they'll do better. But that does not inform future decision making.

a. Reply: See comment #8 above for further discussion of intention to treat analysis and sensitivity analysis for immortal time bias.

b. Changes in text: See changes to limitations in response to Reviewer A.

14) Line 307: I agree with this. This is an interesting aspect of the study but 'patient preference' alone has no detail. In my own practice, i would not expect 40% of patients who underwent neoadjuvant CRT to NOT proceed to surgery and from my knowledge it is not that way in the rest of Europe. It may be something related to the US system. Also, a bigger point has to be made about patient selection.

a. Reply: In results, we discuss patient preference and further stratify patients by documented reason for declining surgery ["Among the 12 patients who declined surgery, 6 (50%) patients cited preference for non-surgical management, 3 (25%) patients cited concern for morbidity, 2 (17%) patients cited advanced age, and 1 (8%) patient cited optimism based on response to chemoradiation. One patient with preference for non-operative management

also cited the lack of a support system to enable post-operative recovery."] This data collection was completed to further characterize the 15% of patients in the cohort that declined surgery. There are many inherent differences between the US and Europe, including primary and preventative care, up to and including access to screening tests that may serve as diagnostic studies for malignancy and epidemiology of risk factors, including risk factors known for specific histology. These factors also vary by state within the US. We recognize that our real-world data is most significantly relevant to patients receiving care in the US system, but is overall applicable to most populations with similar distributions of histology. Please see above to comment 12 regarding possible impact of contemporary RTOG 0436.

b. Changes in text: Additional sentence on patient selection added to conclusion. We also added a sentence emphasizing the importance of early consultation with a surgeon, both in the manuscript itself and the conclusions section of the abstract (minor changes made to introduction to keep at 350 words).