Peer Review File

Article Information: https://dx.doi.org/10.21037/jgo-23-210

Reviewer A

The authors are to be commended for an important effort, which is to gain better understanding of how to use Y90 microspheres to achieve the best possible response. However, the investigation in its current form has too many flaws to be ready for publication. The authors are encouraged to address some critical issues because the interventional oncology community needs this information to optimize patients' outcomes.

Reply: The authors have attempted to address the reviewer's questions as discussed below.

In general, the authors approach this manuscript broadly, combining patients who received resin and glass microspheres into one group. It can be assumed that this was done to improve statistical power. Even combined into one group, the numbers are still small and always statistical power will be suspect. Therefore, it is better to present very clear data with appropriate categorization that the readership will learn from, rather than trying to extract a whole lot of statistics that come across as forced and still of doubtful power, while muddling the findings of importance. Given the marked differences in physical features of resin versus glass this investigation will gain credence to the extent that the data is presented after categorization based on device type. Comparing data between devices is sound, mixing the data is not.

Reply: Thanks for the comment, the paper has been reworked and excludes the combination analysis.

Introduction:
1. Much of the substance of this paper is based on physical characteristics of the available Y90 microspheres. The authors should set the stage in the introduction, particularly regrading the differences in physical characteristics that include greater or lesser specific activity, greater or lesser number of particles for a given amount of radioactivity, and the clear difference in volume of particles given. This is well known to all users of Y90 products and is a surprising omission. This reviewer would like to see in the introduction a sentence of two that establishes the clear difference in specific activity of resin versus glass, and how it is different depending on the number of days of decay from calibration (especially for the glass product). The specific activity of glass in the first part of the first week of decay is far greater than later in the decay curve, which explains major differences in particle number and particle volume delivered by glass v. resin and glass early v. glass late decay. Moreover, resin spheres are in average larger than glass spheres and this also impacts the volume of the particles administered for the same amount of activity. At calibration, a 3 GBq vial of resin is about 44 million particles, and for glass about 1.2 million, therefore resin delivers about 35X the number of particles for the same vial size. This is explained by the very different specific activity. Consider mentioning resin

Flex dose and how it attempts to reduce the difference in specific activity compared to glass.

Reply: Thanks for the comments. The introduction has been reworked with the reviewer's comments in mind.

Materials and Methods:
1. Primary endpoint was radiologic response RR at 3 months. It has been known for a long time that RECIST criteria are inadequate for image-guided loco-regional therapies. This is the reason for development of the m-RECIST criteria, which is now standard for trans-arterial therapies of liver tumors. The data should be re-done using m-RECIST and RECIST omitted. EASL is fine.

Reply: Thanks for the comments. The RECIST criteria has been moved to supplemental data. While the authors whole heartedly agree that it is of significantly less value than mRECIST or EASL in the setting of locoregional therapy, it is still widely applied in the medical oncology data so for that reason was felt to possibly be of some interest. Given the small cohort all patients were found to have the same response classifications when using mRECIST and EASL. Therefore, mRECIST was not added as the authors felt this would be redundant.

2. "glass microspheres were delivered on Tuesdays (n=9(9/17), 52.9%), Wednesdays (n=1 (1/17), 5.9%), Thursdays (n=3 (3/17), 17.6%), and Fridays (n=4 (4/17), 23.5%)". The authors should re-state, in reference to glass microspheres, the days of decay and not day of the week. Any given day of the week could be first-week or second-week decay, therefore they should use days of decay from calibration (Tuesday would be either second or ninth-day decay).

Reply: Thanks for the comments. The requested changes have been made.

3. Nowhere in the Methods is the selection process explained. How was the device type allocated to a particular patient? Is this historical (starting with one product and then migrating to another) or was it based on a set of criteria? There is no table that shows the proportion of lobar versus sub-lobar treatments broken down by sphere type. The reader craves an understanding of the authors' selection bias. Bias is OK in retrospective studies, just tell us about it. It is suggested that table 1 be re-drawn so the bottom half, starting at "microsphere utilized" is presented with resin and glass in separate columns.

Reply: Thanks for the comment. Table 1 has been updated per the reviewers recommendations. The microsphere selection was at the discretion of the performing physician however there did seem to be a tendency to use resin more early and glass more late in the study period, this is discussed in the TARE technique portion of the materials and methods.

Results:

1. It is difficult to know where to start with the interpretation of the ROC analysis. This reviewer suspects that the statistician is facing problems with the small sample sizes and finding curves

that are flat or wander above and below the line of statistical chance. This would explain the perplexing finding of a cutoff point <103 Gy---does this mean that the lower the radiation dose, the better the response? Surely it does not. This is a meaningless finding based on ROC curve that is not statistically strong enough. Better to acknowledge that good ROC analysis is not possible than to present information that reflects the weakness of the data.

Reply: Thanks for the comment. The 103 Gy mark was poorly reported. As the reviewer points out the higher radiation dose was more likely to result in the more favorable response. The authors after consultation with their statistical team did maintain the ROC analysis. While certainly the low numbers is a significant limitation, the Youden's index does provide some insight into the realm of dose that is likely to be needed and the statisticians felt that while certainly not ideal it was a reasonable analysis to apply.

2. It is quite transparent to the reader that the cutoff of >542 Gy is so driven by the glass product that the result for the entire cohort that includes resin is the same >542 Gy! It is quite clear that the data needs to be separated by device type.

Reply: Thanks for the comment, the portion of the paper which combined the two products in analysis were removed.

3. If the ROC curves are pretty much flat and the statistics just not strong enough, then trying to find a cutoff with the Youden index may simply be inappropriate. Please discuss with your statistician.

Reply: Thanks for the comment. The authors after consultation with their statistical team did maintain the ROC analysis. While certainly the low numbers is a significant limitation, the Youden's index does provide some insight into the realm of dose that is likely to be needed and the statisticians felt that while certainly not ideal it was a reasonable analysis to apply.

4. Table 2 shows that 33% of resin cases had either PD or SD and 6% of glass had SD (no PD). This large discrepancy of 33% v. 6% needs to be addressed and statistics applied to the extent that the very small samples sizes allow. These are categorical data that can be compared between two groups, possibly using Fisher exact test or Chi square, whatever applies better. It is puzzling why the authors did not explore this stark difference between the groups. It may be a very fertile ground to analyze statistically.

Reply: This analysis has been added to the beginning of the results section.

5. In the subsection "RR and relationship to dose" it is quite clear that the dose was greater for the glass group compared to the resin group. Analysis requires separation according to the type of sphere used.

Reply: The requested changes have been made and the entire cohort analysis removed.

6. “RR and relationship to particle-load and Specific activity:” Previously, the authors analyzed data based on device type, but in this section no such effort was presented, even though the most fundamental differences between glass and resin are related to particle size, specific activity and the particle load typically given at the time of therapy. It is imperative that the authors present this data categorized by device type because it simply is not scientifically sound to mix devices that are physically so different.

Reply: The requested changes have been made and table 3 has been reworked to reflect device specific data.

7. Once the authors have presented data using m-RECIST and not RECIST, it will be interesting to see if the multivariate analysis changes using the m-RECIST findings.

Reply: When the responses were re-reviewed all patient categorical responses were the same when using either mRECIST or EASL, therefore mRECIST was not included as it was felt to be redundant. However, RECIST has been moved to supplemental material as requested.

8. Table 1 mixes data for both resin and glass regarding activity delivered, particle load and activity per bead. This must be broken down by device type. They cannot be combined given their differences. The term “activity per bead” is not used in other parts of the manuscript. Change to “specific activity”. It would be OK to clarify that specific activity means “activity per bead” in the Methods section. Likewise, Figure 3B should be labeled “specific activity” and not “activity per sphere” to be consistent.

Reply: Thanks for the comment. The requested changes have been made to this table. Figure 3b has been removed based on this and the other reviewers comments.

9. Table 2. Convert to mRECIST.

Reply: When the responses were re-reviewed all patient categorical responses were the same when using either mRECIST or EASL, therefore mRECIST was not included as it was felt to be redundant. However, RECIST has been moved to supplemental material as requested.

10. Table 3. The entire table must be re-worked to account for the physical differences between resin and glass. Can’t mix the two. Must re-draw using mRECIST instead of RECIST, and all data should be categorized by resin and glass.

Reply: When the responses were re-reviewed all patient categorical responses were the same when using either mRECIST or EASL, therefore mRECIST was not included as it was felt to be redundant. However, RECIST has been moved to supplemental material as requested. Furthermore, the table has been reworked to only present data by microsphere type use.

11. Table 4. If the authors believe that meaningful statistics cannot be drawn if the data is divided in two parts to account for device type, this must be presented and discussed. Table 4

in its present form loses meaning insofar tumor dose, specific activity and particle load are concerned.

Reply: Thanks for the comment, the authors humbly disagree. The devices are different, however, becoming more similar all the time as resin provides hotter and hotter flex dose options and glass continues to offer 2nd week dosing. This is not represented in this data set as no flex dosing or 2nd week dose was used. However, in leaving these together it may inform either resin or glass users as to which of these various options may be ideal. This approach also allowed the team to evaluate the effect microsphere type had on outcomes, helping to address a previous comment. Furthermore, the reviewer is correct in their assertion that analysis in this manner after dividing by microsphere type would also not be sound, per our statistician.

12. Figures 1, 2 and 3 have in common a fundamental problem. The cutoffs presented for absorbed dose in Gy cannot be analyzed with the devices mixed. This must be done after separating device used and determining cutoffs that apply accordingly. Same for specific activity. This sub-analysis is particularly important because it is possible that the resin product is more effective at lower radiation dose because of the greater particle load. How will the authors discern these issues if analysis is not done by device type? Additionally, so many KM curves come across as forced into the analysis. It is likely that a simple table showing the responses, survival, and TTP of resin versus glass would be a better stating point that would allow choice of which KM curve to include in the manuscript.

Reply: Thanks for the comment figures, 1 and 2 have been removed.

13. Supplemental Figure 1. Again, the devices are mixed. Unless categorized by device, the authors may not be able to glean important findings that apply to a particular device because of the mixed data. Moreover, the scatter plot charts may ultimately add little and possibly could be omitted.

Reply: Thanks for the comment, the supplemental figures have been removed from the revised manuscript.

Discussion:

Much of the discussion may need to be re-written once the data is analyzed as suggested above. Different issues may come up and different conclusions reached.

Reviewer B

The authors presented a rather small cohort of unresectable iCCA treated with Y90 radioembolization over 10 years at a single institution, aiming to identify the relationship between tumor-absorbed dose and radiologic response as measured by EASL and RECIST criteria. Using iCCA has gained popularity over the last 10 years in the interventional oncology

field, and this topic is of great importance as knowing the threshold of tumorcidal Y90 dose would allow effective treatment planning. However, there are few major weaknesses in this study:
1. Authors should describe tumor characteristics in more detail: tumor number, vascular invasion, and extrahepatic disease.

Reply: Thanks for the comment. The requested details have been added to the first paragraph of the materials and methods.

2. More details are needed regarding patient demographics, such as prior systemic treatment, concurrent treatment, post-TARE treatment, which affect radiologic responses as well.

Reply: Thanks for the comment the requested changes have been made to table 1.

3. Authors should also report adverse events, as high Y90 dose not only leads to tumor necrosis but also hepatic failure. If a high dose Y90 may result in high toxicity or post-chemotherapy delay/withhold, then the use of high dose Y90 can be limited.

Reply: Thanks for the comments, an adverse events portion of the paper has been added the results section of the revised manuscript and is discussed in the revised discussion section.

4. Was pathologic data available in any patients? Did any patient receive resection or transplant?

Reply: No patients were taken to surgical resection or transplanted after treatment, this has been added to the revised manuscript.

Of note, pages are not numbered.
Introduction:

1) "the results of chemotherapy alone are less than desirable" This trial is outdated. Please also include TOPAZ-1.

   Reply: The requested changes have been made to reference 2 of the revised manuscript.

2) "of some locoregional therapy". Too colloquial.

   Reply: Thanks for the comment, the sentence has been reworked with the reviewers comments in mind.

Method:

3) "If patients had multiple TARE treatments of a single-lesion each was included and analyzed separately for the primary endpoint provided the treatments were separated by at least 3-months"—please revise/rephrase.

   Reply: Thanks for the comment, the sentence has been reworked in the revised manuscript.

Results:

4) PD (n=2 (2/26, 7.7%)) or SD (n=2 (2/26, 7.7%), PR (n=13 (13/26, 50%)--suggest format more succinctly, ie PD (2/26, 7.7%). Same for the rest of numbers throughout the manuscript.

   Reply: Thanks for the comment. The suggested changes have been made to the written manuscript and tables.

5) "A ROC curve analysis evaluating tumor dose in those who did and did not achieve an ORR by EASL was performed for the entire cohort and resin only cohort, however, could not be performed for the glass only cohort as only a single patient did not have an ORR"--please revise

   Reply: Thanks for the comment. The sentence has been reworked as requested.

6) "monstrates scatter plots for those who did and didn't achieve a"--please spell out.

   Reply: The authors apologize, but we are unsure what the reviewer is asking to be spelled out. No acronym is utilized in the referenced sentence.

7) "and p=0.65 Cox HR:0.65 (95%CI:0.1- 160 4.16), respectively)."--please revise parentheses.

   Reply: Thanks for the comment the requested changes have been made.

8) "Figure 2 demonstrates KM curves for TTP, local-TTP, and OS comparing those who did and did not receive ≥294 Gy to the tumor. The TTP (p=0.94, Cox HR:0.96 (95% CI: 0.35-2.63)) and local TTP (p=0.81, Cox HR:0.82 (95% CI: 0.16-4.15)) did not show significant differences." – What are the logrank test? Does data fulfill criteria for Cox proportional model?

   Reply: Based on the other reviewer comments these figures have been removed.

Discussion:

", especially in conjunction with systemic chemotherapy further data is needed (8-17)"—please revise

Reply: The sentence has been reworked with the reviewers comments in mind.

"In particular more data on dose thresholds is needed"—authors can include data on dose/effectiveness on HCC to show why knowing dose thresholds on iCCA is important.

Reply: This sentence has been reworked with the reviewers comments in mind.

“Similarly, if you maintain a constant total activity delivered but drop the activity per particle a higher particle load will be necessary, and this may affect factors such as TNR”--too colloquial. Suggest authors to improve scientific writing throughout the manuscript.

Reply: Thanks for the comment, the sentence has been removed.

Tables:
Table 1: Delivery Target: any modified lobectomy?

Reply: Thanks for the comment, no modified lobectomy technique was used.

Table 2: Suggest list as count/total, percentage. I.e in the first column: “2/26, 7.7% instead of 2(2/26, 7.7%)”

Reply: The requested changes have been made to all tables and written manuscript

Table 4: Please list OR, 95CI, and p values in separate columns.

Reply: Respectfully the authors felt that making these changes made the table more difficult to follow. The suggested changes are shown below, but not in the revised manuscript. If the editor feels strongly that the suggested format is superior then the authors are willing to make the changes.

Variable	Univariate regression OR:(95%CI),*p* value			Multivariate regression OR:(95%CI),*p* value		
	OR	95%CI	*p* value	OR	95%CI	*p* value
Tumor dose	1.26	1.08-1.58	0.017	1.22	0.99-1.61	0.28
Specific activity	1.16	1.01-1.31	0.045	1.01	0.98-1.37	0.48
Perfused volume	0.99	0.94-1.16	0.28	-	-	-
Particle load	1	0.98-1.03	0.74	-	-	-
Tumor size	0.75	0.53-1.05	0.086	0.94	0.61-1.45	0.79
Pretreatment CA 19-9	0.97	0.94-1.45	0.18	-	-	-
Ytrrium-90 material (resin or glass)	0.14	0.01-1.38	0.07	2.99	0.05-3.89	0.61
Prior chemotherapy	7.3	1.46-14.71	0.008	6.88	0.49-15.40	0.15