Peer Review File

Article Information: https://dx.doi.org/10.21037/jgo-21-207

Reviewer A

In this NCDB study, the authors focus their analysis on Hispanic patients to identify variables independently associated with presentation, treatment, and outcome. The introduction nicely summarizes the potential for biological differences among different racial, ethnic, or geographical populations. The authors focus the analysis on Hispanic and non-Hispanic white patients, which is unique. The findings are certainly interesting as other studies have focused on the aggressive nature of malignancy in younger Hispanic populations.

Comments: Can the authors provide further information or a reference on how collaborative stage site-specific factor 25 excludes Siewert type I and II tumors?

Reply: For patients who are staged by AJCC 7th edition or earlier, collaborative stage site-specific factor 25 records the distance of the tumor midpoints from the EGJ, which allowed us to exclude all Siewart I-III tumors for all patients and exclude patients who would now be managed as esophageal cancer.
Changes in text: this has been clarified in the methods section – page 5, line 101

Page 7, line 132. There may be a typo as 54% of patients were 18-40 years old and 42% were >70 years old. What about those age 41 to 70?
Reply: Thank you for identifying this error: The majority of patients were ages 41-70 (53.9%) and ages 70+ (42.3%), while ages 18-40 represented 3.8% of the study population.
Changes in text: this has been edited in the results section and tables – page 7, line 150

What were the case #'s associated with the definition of high and low volume cohorts? In the conclusion, the authors emphasize the benefit of referral, but it would be helpful to know what annuals case #'s constitute a high volume center.
Reply: The <50th, 50-75th, and >75th percentile correspond to <56, 57-109, and >109 cases over 11 years, respectively.
Changes in text: this has been added to the methods section – page 5, line 110

It would be also helpful to know the racial classifications and numbers of patients that were excluded from the study. Non-Hispanic white is a very large & broad category (approximately 78K) for analysis. Does the NCDB classify middle eastern Americans as white? Are there any other racial or ethnic groups that are included in the white category, and could this be a limitation?
Reply: The NCDB method of defining race and ethnicity follows the 1997 Office of Management and Budget (OMB) standards on race and ethnicity that the US Census Bureau uses as well. It clearly separates out patients who define their origins as Black or African

American, American Indian, Asian, and Native Hawaiian or Other Pacific Islander by self-identification. Per the OMB standards, White includes "a person having origins in any of the original peoples of Europe, the Middle East, or North Africa." We agree that both Hispanic and non-Hispanic white groups are highly heterogeneous and is a limitation to our findings.
Changes in text: This has been clarified in the methods and limitations section – page 5, line 96

Reviewer B

The burden of gastric cancer involving Hispanic patients in the United States is growing as both the population and the incidence of gastric cancer in this group increase. I would recommend to state as both the population and incidence of gastric cancer in this group is increasing
Reply: Thank you for improving the grammar and terminology of our manuscript
Changes in text: this has been changed in the abstract, page 2, line 38

Reviewer C

Methods:
- It would be helpful to know which specific ICD-O-3 morphology codes were captured
Reply: ICD-O-3 morphological codes have been included in the methods section and re-classified to signet vs other, which was the main comparison throughout the study.
Changes in text: page 5, line 94

- Please more clearly state that only patients with Caucasian race were selected for analysis, and explain why other races were excluded
Reply: Race and ethnicity were assessed via the same schema as the 1997 Office of Management and Budget standards used by the US Census Bureau, in which white race includes a person having origins in any of the original peoples of Europe, the Middle East, or North Africa. The primary goal of our manuscript was to evaluate the experience of Hispanic patients with gastric cancer. We believe that this population is important to study as they tend to have aggressive disease at a younger age, and yet have improved survival. We chose the predominant population in the US for the comparison group.
Changes in text: Selection criteria for white race has been clarified in the methods section – page 5, line 96

- Given that the majority of patients resided in metropolitan areas, providing a more granular assessment of distance traveled seems appropriate i.e. <5 miles or <10 miles instead of <25 miles
Reply: Thank you for the excellent advice. Distance traveled has now been changed to <10, 10-25, and >25 throughout the study. The Cox regression was re-done to reflect these changes.
Changes in text: Methods: Page 5, line 109

- Tumor location in the NCDB is provided by ICD-O-3 topographical codes, not ICD-9 diagnosis or procedure codes

Reply: “ICD-9” has now been replaced ICD-O-3 topographical codes in the methods section.
Changes in text: Page 6, line 116

Results:
- The bimodal age distribution seems odd to characterize gastric cancer and is likely incorrect; a quick assessment of NCDB data from 2004 to 2017 for adenocarcinoma and signet ring cell carcinoma (ICD-O-3 codes 8140 and 8490, representing 150,000 patients) reveals that <5% of patients are age <40, 50% are age 41 to 70, and 47% are age >70; even the application of additional selection criteria will not substantially change that distribution enough to match yours
Reply: Age categories were incorrectly labeled in the table and the results section. The majority of patients were ages 41-70 (53.9%) and ages 70+ (42.3%), while ages 18-40 represented 3.8% of the study population.
Changes in text: page 7, line 150, and in the tables

- It should be made much more clear to the reader in the Results (and discussed in the Methods) that Table 2 and Figure 1A-1D are constructed based on analytic stage, which is a combination of pathologic staging when available and clinical staging when pathologic staging is not available; given the high degree of variability in pathologic and clinical staging in gastric cancer, this is not a trivial point to make
Reply: This is an excellent point and has been corrected in the tables and manuscript
Changes in text: page 6, line 119 and in the tables

- Cox is a proper name should be capitalized as such
- One is left to assume that the Cox proportional hazards model as described in the text and depicted in Table 3 is multivariable, as the authors use the word 'independent' to describe prognostic factors, but this is not made explicitly clear either in the Methods or the Results
Reply: Cox regression is now appropriately capitalized, and is now clearly stated as multivariable in the methods section.
Changes in text: Page 6, line 137

- In Table 3, it would be helpful to state the raw number of patients for each subcategory in a column to the left of the Hazards Ratio; 'Case Volume' data is repeated; it's Charlson Deyo not Dayo
- In Table 3, did you mean to say staging was all pathologic staging, or was it analytic stage?
- In Table 4, should it not be pathologic stage, instead of analytic stage?
- In Table 4, what is the threshold for 'Yes' for Lymphadenectomy? At least one node harvested or at least 15 nodes harvested?
- In Table 5, as with Table 3, it would be helpful to state the raw number of patients for each subcategory in a column to the left of the Hazards Ratio; decide if you will hyphenate Charlson-Deyo or not and keep it consistent throughout
- Why was the LN harvest cutoff 16 nodes instead of 15 in the first paragraph of the 'Surgical Cohort' section?
Reply: The tables have been edited to remove case volume. Charlson-Deyo is now consistently

described as such. In addition, staging is now clearly labeled as analytic stage. Lymphadenectomy has now been clarified as "⩾16 lymph nodes harvested", and staging is now clearly labeled as analytic stage
Changes in text: Table 3, 4 and 5; Page 9, line 201

Discussion:
- Generally, all results should be reported in the Results for the first time; true 'incidence' cannot be commented on using the NCDB, as it is a national cohort and not a true population registry like SEER is; therefore, the most you can comment on is the relative proportion of Hispanic patients across the years of study, which again should be first mentioned in the introductory paragraph of the Results
Reply: Thank you for the clarification – the results section no longer describes "incidence" and now describes the rate of growth over time.
Changes in text: page 7, line 150

- A significant limitation of the NCDB is its selection bias; in 2010, the US population was approximately 16% Hispanic, but since the NCDB only captures patients treated at CoC centers, there is an ethnic/racial bias to its patient selection
Reply: We agree that selection bias in terms of race and ethnicity is a significant limitation to our study. This has been added to the limitations section.
Changes in text: page 15, line 320

Overall:
- This study presents a great deal of data, but a more refined research question would improve its impact; I think it would be worthwhile to consider focusing on surgical patients and enhancing the analysis by a propensity score analysis to compare white Hispanic to non-Hispanic patients in order to control for the clinicopathologic disease factors you have available to you in the NCDB, thereby actually furthering the understanding of why Hispanic patients seem to fare better than their non-Hispanic white counterparts.

Reply: thank you for the thorough review of our manuscript. We performed 1:1 propensity score matching for the surgical cohort to balance the covariates available to us in the NCDB. It has now been added to the methods, results, discussion and tables.

Reviewer D

Tseng et al have used the NCDB to evaluate the presentation, management, and outcomes of Hispanic patients in the United States who are diagnosed with gastric cancer. These patients are compared to non-Hispanic white patients. Despite being socioeconomically disparate, having unfavorable histopathologic features, and less likely to receive adjunct therapies, Hispanic patients had better overall survival.
The paper is well written, the conclusions are not overstated, and the limitations are addressed. I have the following comments/questions:

On page 9, line 178 it states that Hispanic patients were more likely to have >15 LN evaluated (and again in a few other places), but in the paragraph above, the current staging recommendation of ≥ 16LNs is referenced. This needs to be reconciled and consistent through the manuscript.

Reply: The manuscript now consistently uses the term "⩾16 lymph nodes" instead of ">15 lymph nodes"

Changes in text: Page 9, line 201

Are the authors aware of any literature from Mexico or Latin America regarding the incidence and behavior of gastric cancer and how that compares to the results in this study – similar to the comparisons made between patients in East Asia and Asian Americans? It would be worth including in the discussion for context – if there is data available.

Reply: Thank you for the excellent suggestion. The majority of articles regarding tumor biology in the Hispanic population are based in the United States. There are very few articles regarding the experience of Hispanic gastric cancer patients outside of the US. In our literature review, we found a meta-analysis of 29 retrospective, small sample size studies in which 9 of the studies described tumor characteristics (location and histology) – this has been included in the discussion. In addition, we found an abstract that describes gastric cancer in Mexico, which found an increasing incidence of young gastric cancer patients. They also found a higher rate of signet histology and a higher rate of metastatic disease on presentation compared to our findings. This has been added to the discussion.

Changes in text: Page 13, line 281