Peer Review File

Article information: https://dx.doi.org/10.21037/gs-23-258

Reviewer A

This is an article about an interesting analysis of the SEER database regarding lymph node metastasis.
Please consider the following suggestions for improvement:
*) I would use a lowercase for "grade" instead of a capital letter ("Grade"), e.g. Abstract, Results section, line 42.

Response: Thank you for your recommendations. We have incorporated revisions to correct the improper use of uppercase letters as per your guidance. The corrections have been implemented in the following sections: Line 42, 45, 49, 140, 151, 152, 183-184, 194, 272, 497.

*) Reference 3 is not suitable in the Introduction section, line 62: "… and increased risk of disease recurrence …", because this review (reference 3) is about oesophageal cancer. It should be substituted.

Response: Thank you for your meticulous review. We have replaced the reference "Extranodal extension of lymph node metastasis is a marker of poor prognosis in oesophageal cancer" from reference 3 with the reference "Extra-nodal extension of sentinel lymph node metastasis is a marker of poor prognosis in breast cancer patients: A systematic review and an exploratory meta-analysis" (See References section, Line 537-541).

*) Results section, 3.2, line 150: "age" instead of "Age"

Response: Thank you for your response. We have updated the term "Age" to "age" in line 150 and made corresponding adjustments in similar instances (Line 140, 141, 165).

*) Discussion section, line 255: Figure 1?

Response: Thank you once again for bringing this to our attention. We have rectified the error where we incorrectly referenced "Figure 1" when it should have been "Figure 2." The correction has been made in line 277.

*) Discussion section, lines 412/413: "The Chinese text you provided appears to be discussing the results and considerations of neoadjuvant therapy in Luminal B-like breast cancer." ???

Response: Thank you for addressing this matter. We appreciate your clarification. The

sentence "The Chinese text you provided appears to be discussing the results and considerations of neoadjuvant therapy in Luminal B-like breast cancer" has been removed, and we have made the necessary contextual modifications in lines 350-358.

*) Discussion section, line 420: “NACT” instead of “NACTT”

Response: Thank you for the correction. We have now rectified "NACTT" to "NACT" in line 461.

*) Figure 1: please insert a comma in e.g. 256,504 and 23,774, uniformly.

Response: Thanks for your advice. We have made the necessary corrections accordingly. The formal versions of the numbers have been applied to Figure 1, Figure 2, Figure 5, Figure 6, Table 1, Table 2A, Table 2B, and Supplement 1A-1G.

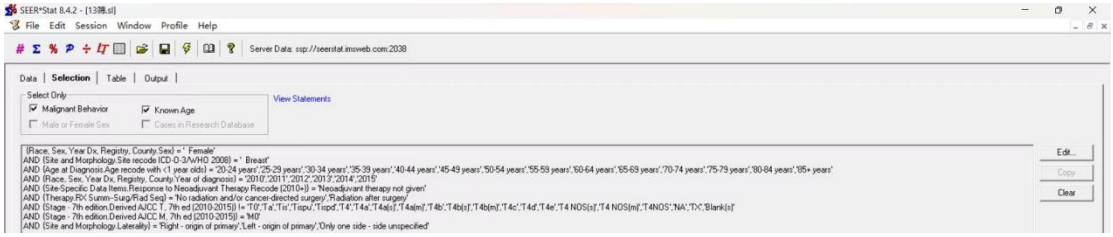
Reviewer B

The study is a proposal with clinical relevance, despite being a relatively exhausted approach (nomograms for lymph node metastasis). Despite this, the manuscript has many weaknesses:
Grammar must be professionally corrected. There are several very long sentences and others with incorrect meanings (see line 58 sentence).

Response: Apologies for any confusion caused by our initial writing. We have now thoroughly revised and polished our manuscript, rectifying the identified grammatical errors. The revised sentences have been clearly indicated with red markings (Lines 60-63).

In Materials and Methods, the authors mention data collection from 256,504, and later there were more exclusions. Also, the structure of the article is poor. Authors should check and adhere to the STROBE guidelines in the way of reporting the observational study, also improving the inclusion/exclusion flowchart by variables.

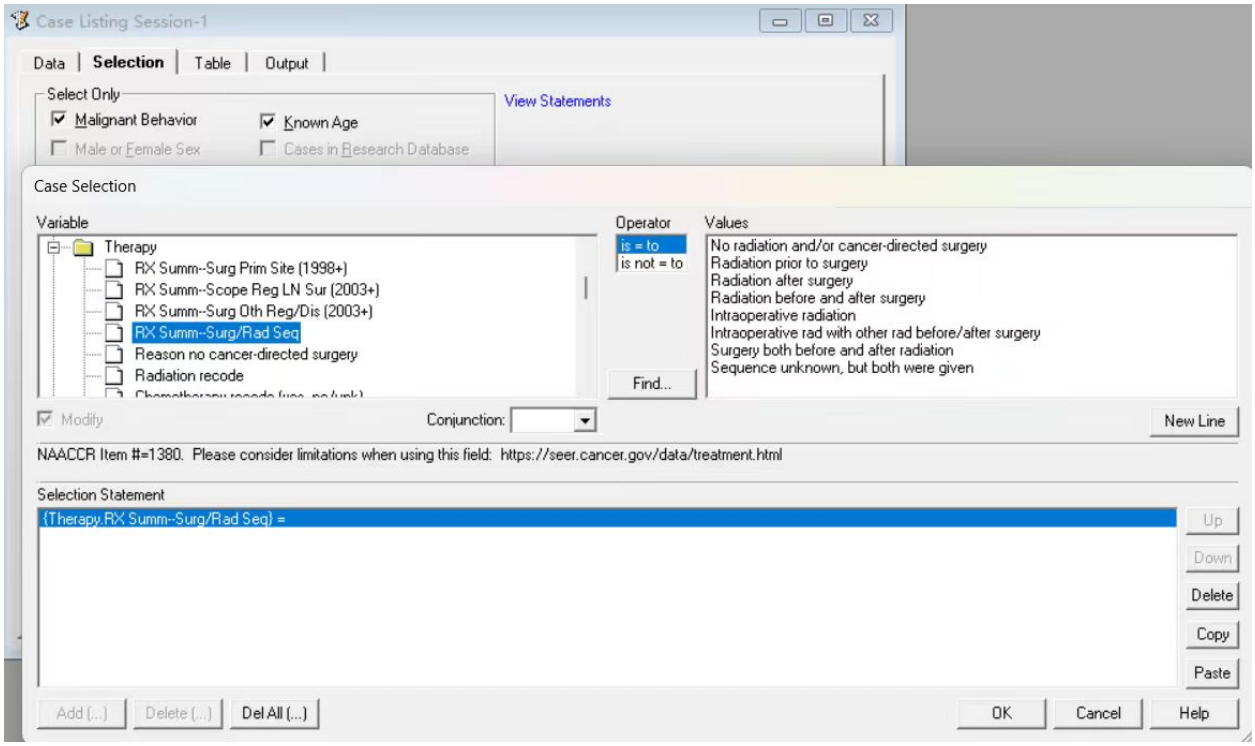
Response: Thank you for your guidance. Regarding the discussion on enhancing the process flowchart for inclusion/exclusion, we originally gathered a dataset consisting of 256,504 cases using the specified inclusion and exclusion criteria (refer to the figure below). Subsequently, we simplified this process by implementing a single step, which involved the exclusion of 23,774 patients who did not undergo lymph node dissection. In response to your valuable suggestion, we have now revised the flowchart accordingly (see Figure 1).



Lines 83-84: "No radiation and/or cancer-directed surgery & Radiation after surgery". Wouldn't it be radiation before surgery? Why exclude patients who received adjuvant radiotherapy?

Response: Thank you for your professional feedback. As illustrated in the figure below, which outlines the data selection and screening process in Therapy RX Summ--Surg/Rad Seq, there are eight available options: "No radiation and/or cancer-directed surgery," "Radiation prior to surgery," "Radiation after surgery," "Radiation before and after surgery," "Intraoperative radiation," "Intraoperative rad with other rad before/after surgery," "Surgery both before and after radiation," and "Sequence unknown, but both were given."

To ensure that the factors influencing lymph node positivity rate do not interfere with our study, we have excluded any potential variables that may do so. Therefore, we have chosen to focus solely on patients falling within the "No radiation and/or cancer-directed surgery" and "Radiation after surgery" categories. This decision was made because interventions within these two categories do not impact the detection of involved lymph nodes, aligning with the primary focus of our study on clinicopathological characteristics associated with lymph node metastasis.



While I agree with grouping ethnicities and histologies into the "others" group/category, the subtype cannot be classified as "others" unless it is an unknown subtype.

Response: Thank you for your correction. We have made the adjustment to replace the "others" category with "unknown" in the figures, tables, and nomogram.

Also in Materials and Methods, the authors mention survival analyses, and such results are never presented, nor are they in harmony with the proposed objective.

Response: Thank you for bringing our mistake to our attention. We acknowledge that this error was a result of our oversight. It should be noted that we did not include survival analysis in our study. Consequently, we have taken the necessary steps to remove all content related to survival analysis from the manuscript. (Line 125-126, 137-138).

Also, the PSM is a technique for pairing patients with regard to potential confounders of an exposure variable (treatment, etc.), even more so if it involves the outcome (definition of confounding variable). This critical point of the study was not described, and I do not see any purpose, since no outcome was analyzed.

Response: Thank you for your comments. We conducted univariate and multivariate logistic regression analyses, taking into account the outcomes of positive and negative lymph nodes. Furthermore, we performed propensity score matching (PSM) for each factor as presented in Table 1. These analyses were undertaken to explore potential

correlations between different clinical and pathological characteristics and lymph node metastasis.

In response to your valuable recommendations, we have now provided a clear description of the purpose and significance of PSM within the literature, and we have also presented a thorough analysis of the results. (Line 77-86, 263-269).

In the Discussion, the authors present new results and digress to NACT. The relationship of NACT with the proposed objective is understood, but the authors did not carry out any type of analysis with NACT to discuss so much about this subject.

Response: Thank you for your insightful correction. We acknowledge that we may have overly emphasized neoadjuvant therapy in this section without adequate research support. To address this, we have eliminated redundant content and extensively revised the relevant portions (Lines 359-476, 350-358).

As limitations, the authors mention the absence of data on ER/PR and HER2. However, the SEER database makes them available qualitatively (see links: https://web2.facs.org/cstage0205/breast/Breastschema.html; https://staging.seer.cancer.gov/eod_public/schema/ 3.0/breast/).

Response: When discussing the limitations of this article, you kindly provided us with a valuable website for reference. We have thoroughly explored the provided information and deeply appreciate your extensive knowledge. We believe that this resource will significantly enhance our future research efforts. We do, however, regret the limitation in our initial study. We mistakenly stated that “The database does not include additional information such as Ki-67, ER/PR, and HER2+ component content.” Actually, the information related to ER, PR, HER2 expression were available in SEERdata. We have re-written this part as “The database does not include additional information such as Ki-67, vascular embolism, lymphovascular invasion, nerve invasion, etc…” (Lines 479-481)

The website you recommended includes laboratory values for HER2 immunohistochemistry (IHC) via CS Site-Specific Factor 8. In our upcoming research, we intend to investigate the impact of varying levels of ER/PR/HER2 on tumor metastasis to lymph nodes and others. Given that breast cancer molecular subtyping in clinical practice primarily falls into five categories: HER2-positive, HER2-negative, triple-negative, luminal A, and luminal B, we have decided not to separately present this molecular subtyping in a nomogram within this manuscript. Instead, we plan to analyze the effects of different receptor component levels in tumor tissue on tumor cell invasion and metastasis through clinical data analysis in our future work. Your input has been highly valuable, and we appreciate it greatly.

The authors talk a lot about the differences depending on the number of resected lymph nodes. In fact, the number of lymph nodes analyzed may increase the number of affected lymph nodes observed, as described in the vast literature. Although not applicable to a nomogram, models with a count-dependent variable (negative Poisson/Binomial) can be adjusted for such exposure. Authors could develop such models to help balance for this difference.

Response: Thanks for your insightful comments, which is very constructive. In our current study, we are focusing on contruct a nomogram model for predicting axillary LN involvement. Your suggestion is very valuable, you offered a constructive suggestion to utilize models with a count-dependent variable (negative Poisson/Binomial) for analysis and research. Although we are not yet proficient in this method, we will make an effort to further explore it in the future.

Regarding lymph node metastasis and its association with the subtype being apparently contradictory, the literature shows that hormone receptors regulate the CXCL12/CXCR4 axis, which favors both lymph node and bone metastasis. There is also blood vessel metastasis, and the literature shows that approximately 25% of distant metastases originate from lymph node metastasis (DOI: 10.1016/j.ebiom.2020.102793).

Response: We sincerely appreciate your extensive knowledge and valuable insights. Through an in-depth study of your perspectives and the literature you provided, we have made revisions to pertinent sections of our paper, further enhancing the depth of our research. Triple-negative breast cancer is known for its poor prognosis, yet both our research and that of others suggest a relatively lower rate of axillary lymph node metastasis in this subtype, which appears contradictory. Upon careful examination of the literature you supplied (as cited in the discussion section), we have come to understand that breast cancer can progress to distant metastasis without involving axillary lymph node metastasis as an intermediate step. Consequently, the lower rate of lymph node metastasis in triple-negative breast cancer may be associated with a higher probability of distant metastasis, potentially impacting its prognosis. (Lines 329-343)

Reviewer C

Well done study. It is true that it would be much more interesting if it included Ki67 and lymphovascular invasion, but it is well organized and the results are consistent. In my databases, triple negative tumors also have less involvement of the sentinel node.

Response: Thank you for your valuable comments. Unfortunately, the SEER database does not provide data on Ki67 and lymphovascular invasion. Nevertheless, your suggestion has provided us with valuable insights, and we intend to collect this data from clinical patients at our own hospital for in-depth analysis. We are pleased to learn that the data generated from our SEER dataset aligns with your findings. This reaffirms

the validity and reproducibility of the results obtained in our current study.

Reviewer D

I would like to highlight some positive aspects of your work that led to its acceptance:

Clear and Objective Writing: Your article is exceptionally well-written, with a clear and concise presentation of the research methodology and findings. The clarity of the language used greatly enhances the readability and comprehension of the study.

Precise Conclusions: The conclusions drawn from your research are precise and supported by the data presented. Your study's results contribute significantly to the understanding of the topic.

Objective Tables and Figures: The tables and figures provided in the manuscript are well-designed and effectively illustrate the data, making it easier for readers to interpret the results.

I believe your research adds valuable insights to the existing literature and addresses an essential aspect within the field.

In conclusion, I have not suggested alteration or revising the text.

Response: Thank you for your review work and positive comments.