**Peer Review File**

Article information: https://dx.doi.org/10.21037/jgo-23-125

**Reviewer A**

First of all, my major concern for this study is that the current data are not feasible to answer the research question of the predictive accuracy of the prediction model. Due to the lower AUC values of the model in both training and validation samples, the study part on the predictive accuracy has been failed. I suggest the authors to focus on the risk factors only for this study.

Reply: We thanks for your suggestion. We have revised the discussion and conclusion accordingly. Moreover, we have discussed the limitation of the low predictive accuracy of the prediction model in the discussion. See page 7, line 217-225.

Second, the title needs to indicate the clinical research design, i.e., a retrospective cohort study based on the SEER.
Reply 2: We have revised the title according to your suggestion. See page 1, line 2-3.

Third, the abstract needs some revisions. The background needs to indicate the limitations of prior studies and what the knowledge gap is in relation to the research focus of the risk factors of lymph node metastasis. The methods need to describe the inclusion of subjects, the diagnoses of CRC and lymph node metastasis, and assessment of potential risk factors. The results need to briefly describe the clinical characteristics of the study sample and quantify the findings by reporting accurate OR and P values. The conclusion needs comments for the clinical implications of the findings, not to repeat the main findings again.
Reply 3: We have revised the abstract accordingly. However, we can not add some related contents in the abstract as the editor required that the words of the abstract could not be more than 350 words (Now it was 350 words). See page 2, line 35-61.

Fourth, in the introduction of the main text, the authors need to briefly review what has been known on the incidence rates of lymph node metastasis and its associated factors, and have comments on their limitations and knowledge gaps, to support the clinical needs for this research focus.
Reply 4: We have added related contents in the introduction. We thank for your positive feedback. See page 3, line 85-96.

Fifth, in the methodology of the main text, please describe the SEER database in detail, describe the details of the diagnoses of CRC and lymph node metastasis, and definitions

of potential risk factors, as well as the patient follow up. In statistics, please describe the details of the multiple logistic regression in particular how factors were included and selected. The calculation of ORs should be described.
Reply 5: We have revised the methods accordingly. See page 4, line 105-106 and page 4, line 126-130.

**Reviewer B**

The paper titled "Risk factors of lymph node metastasis in patients with T1 stage colorectal cancer based on the Surveillance, Epidemiology, and End Results database" is interesting. Age at diagnosis, rectosigmoid cancer, poorly differentiated or undifferentiated tumor cells, and distant metastasis are independent risk factors of lymph node metastasis in patients with T1 stage CRC. The early identification of risk factors of lymph node metastasis in patients with T1 stage CRC will help clinicians to choose the appropriate treatment measures. However, there are several minor issues that if addressed would significantly improve the manuscript.

1) The content of this study is too simplistic. It is suggested to increase the role of tumor infiltrating lymphocyte in predicting lymph node metastasis in patients with T1 colorectal cancer.
Reply 1: The data of the tumor infiltrating lymphocyte in the SEER database was limited. Therefore, we failed to study the data of the tumor infiltrating lymphocyte. We have discussed this as a limitation in the discussion. See page 7, line 224-225.

2) Compared with other prediction models, what are the advantages of the model in this study? It is recommended to add relevant content.
Reply 2: The number of patients enrolled in a previous study was relatively small. We have added this in the introduction. See page 3, line 85-96.

3) The histopathology and tumor characteristics of tumors associated with lymph node metastasis should be included in the introduction.
Reply 3: These were introduced in the introduction. See page 3, line 85-96.

4) Figures 2-5 are not clear enough. It is recommended to provide clearer figures again.
Reply 4: We have replaced the figures 2-5. See the attached files.

5) How to judge the prognostic characteristics of colorectal cancer based on the results of this study? How to provide candidate targets for the treatment of colorectal cancer? It is recommended to include relevant descriptions in the discussion.
Reply 5: For T1 stage CRC patients with the above-mentioned risk factors, endoscopic physicians should carefully evaluate the advantages and disadvantages of endoscopic surgery before deciding whether to perform this surgery. See page 6, line 200-202.

6) The introduction part of this paper is not comprehensive enough, and the similar papers have not been cited, such as “Developing and validating a multivariable machine learning model for the preoperative prediction of lateral lymph node metastasis of papillary thyroid cancer, Gland Surg, PMID: 36761483”. It is recommended to quote this article.
Reply 6: We have revised the introduction according to your suggestion. And we have inserted the reference. See reference 10. See page 3, line 85-96.

7) How to identify and verify the prognostic characteristics for predicting disease-free survival and overall survival of patients with colorectal cancer by integrating multiple data sets? It is recommended to add relevant content.
Reply 7: This study mainly explored the relevant factors of lymph node metastasis in T1 stage CRC patients, providing a reference for clinical physicians to choose treatment plans. Therefore, there is no in-depth exploration of prognostic factors.

**Reviewer C**

Authors analyzed patients with T1 CRC and a multivariate logistic regression analysis was conducted to analyze the risk factors of lymph node metastasis. They established a prediction model to predict lymph node metastasis. How to treat patients with T1 CRC is controversial and important, but there are some problems that needs to be reconsidered in this article.

1. Multivariate logistic regression analysis was performed and age at diagnosis, rectosigmoid cancer, poorly differentiated or undifferentiated tumor cells, and distant metastasis were independent factors of lymph node metastasis. Why did you include distant metastasis? The possibility of lymph node metastasis will change the decision of surgery with lymph node dissection, but if the patient has distant metastasis, the patient must be Stage IV, it would be a completely different treatment strategy. Please reconsider the meanings of prediction and select appropriate variables.
Reply 1: We included all T1 stage CRC patients regardless of whether they had local or distant metastasis. Multiple factor regression analysis was conducted simultaneously to exclude the mutual influence between different variables.

2. They established the prediction model for the lymph node metastasis, but the AUCs were under 0.7. Commonly, the AUC < 0.7 is not good model. There are several reports of prediction model for the lymph node metastasis in patients with T1 CRC, and their AUCs are more than 0.7. (Ex, Wada et al reported the prediction model with AUC>0.8, Fujino et al reported the model with AUC > 0.7)
Reply 2: We have discussed the low predictive value of the model as a limitation in the discussion. See page 7, line 217-221.

3. Why have we not considered the known risk factors for lymph node metastasis, such as vascular invasion and SM invasion depth? They must be included.
Reply 3: The data of vascular invasion and submucosa invasion depth was not available in the SEER database. We have discussed this as a limitation in the discussion. See page 7, line 224-225.

4.Is the pathological type the main one? As it has recently been reported that the most poorly differentiated of the types included in T1 colorectal cancer is a risk factor. These also need to be considered.
Reply 4: In table 1 and table 2, the data showed that poorly differentiated or undifferentiated tumor cells was a risk factor. See page 13-14, table 1-2.