79487_Auto_Edited.docx

WORD COUNT 3735
TIME SUBMITTED 12-DEC-2022 04:47PM
PAPER ID 93899271

Name of Journal: *World Journal of Gastroenterology*
Manuscript NO: 79487
Manuscript Type: ORIGINAL ARTICLE

Observational Study

Survival benefit of younger gastric cancer patients in China than the United States: a comparative study of survival, prediction model, and biological analysis

younger gastric cancer patients

Penghui Niu, Lulu Zhao, Wanqing Wang, Xiaojie Zhang, Zefeng Li, Xiaoyi Luan, Yingtai Chen

Abstract

BACKGROUND

The impact of racial and regional disparity on younger patients with gastric cancer (GC) remained unclear.

AIM

The study sought to investigate the clinicopathological characteristics, prognostic nomogram, and biological analysis of younger gastric cancer patients between China and the US.

METHODS

From 2000 to 2018, GC patients aged less than 40 years were enrolled from the China National Cancer Center and the Surveillance Epidemiology and End Results (SEER) database. Biological analysis was performed based on the Gene Expression Omnibus (GEO) database. Survival analysis was conducted *via* Kaplan-Meier estimates and Cox proportional hazards models.

RESULTS

Finally, a total of 6098 younger GC patients were selected from 2000 to 2018, of which 1159 were enrolled in China National Cancer Center, 4939 were collected from SEER database. Compare with the US group, younger patients in China revealed better survival outcomes ($P < 0.01$). For race/ethnicity, younger Chinese cases also enjoyed a better prognosis than that in White and Black sets ($P < 0.01$). After stratification by pTNM stage, survival advantage was observed in China with pathological stage I, III, and IV (all $P < 0.01$), whereas younger GC patients with stage II showed no difference ($P = 0.16$). In multivariate analysis, predictors in China involved period of diagnosis, linitis plastica, and pTNM stage, while race, diagnostic period, gender, location, differentiation, linitis plastica, signet ring cell, pTNM stage, surgery, and chemotherapy were confirmed in the US group. Later, prognostic nomograms for younger patients

were established, with the area under the curve (AUC) 0.786 in China group and 0.842 in the US group. Moreover, three gene expression profiles (GSE27342, GSE51105, and GSE38749) were enrolled in further biological analysis, and distinctive molecular characteristics were identified in younger GC patients among different regions.

CONCLUSION

Except for younger cases with pTNM stage II, survival advantage was observed in China group with pathological stage I, II, and IV than the US group, which might be partly due to differences in surgical approaches and the improvement of the cancer screening in China. Nomogram model provided an insightful and applicable tool to evaluate the prognosis of younger patients in China and the US, respectively. Furthermore, biological analysis of younger patients was performed among different regions, which might partly explain the histopathological behaviors and survival disparity in the subpopulations.

Key Words: Gastric cancer; Younger patients; Racial disparity; Regional disparity; Prediction model; Biological analysis.

Niu P, Zhao L, Wang W, Zhang X, Li Z, Luan X, Chen Y. Survival benefit of younger gastric cancer patients in China than the US: a comparative study of survival, prediction model, and biological analysis.. *World J Gastroenterol* 2022; In press

Core Tip: The impact of racial and regional disparity on younger patients with gastric cancer (GC) was not clear. A total of 6098 younger GC patients were selected from 2000 to 2018, of which 1159 were enrolled in China National Cancer Center, 4939 were collected from SEER database. Compare with the US group, younger patients in China revealed better survival outcomes. For race/ethnicity, younger Chinese cases also enjoyed a better prognosis than that in White and Black sets. Later, prognostic nomograms for younger patients were established, with the AUC 0.786 in China group

and 0.842 in the US group. Moreover, three gene expression profiles (GSE27342, GSE51105, and GSE38749) from GEO database were enrolled in further biological analysis, and distinctive molecular characteristics were identified in younger GC patients among different regions.

INTRODUCTION

Gastric cancer (GC) was the fourth leading cause of cancer-related death worldwide [1], with the highest incidence among individuals aged 50-70 years and only 2.0-6.2% documented incidence among patients younger than 40 years [2,3]. Although the incidence of GC has gradually declined, important age-specific details may be obscured by the overall trend [4]. During the last decades, the incidence of younger GC patients has remained stable or even increased in both Western and Eastern populations [5,6,7].

It was long thought that younger patients with GC had aggressive behavior and serious prognosis [8]. Further insight into which subpopulations were at the highest risk of dying from each remained a crucial requisite so that interventions can be initiated appropriately. Recently, several studies have demonstrated a wide survival discrepancy of GC in different regions or races [9,10]. Notably, GC patients in Asia had more favorable prognosis than Western. The important difference has also been found in certain subtypes of GC, including gender and anatomic location [9,10]. For younger GC patients, however, knowledge concerning the regional and racial disparity was scarce. Chen *et al* [11] has reported that younger patients in China own a longer survival time than the US, whereas Strong *et al* [12] showed no difference between the US and Chinese cohorts. The inconsistent findings from these studies might derive from the small sample sizes, with the population records ranging from 336 to 1,075.

As such, based on a unique combination of the SEER database in the US and a high-volume National Cancer Center Database in China, we sought to compare the clinicopathological features, survival outcomes, and prognostic nomograms in younger Chinese and US patients. Moreover, biological analysis of younger GC patients was

further evaluated, which might partly explain the histopathological behaviors and survival disparity among different regions and races.

MATERIALS AND METHODS

Data source and patient selection

The study queried clinical data from 2000 to 2018 based on the 2 Large independent cohorts. The histologically confirmed GC cases in China were selected through the China National Cancer Center Gastric Cancer Database (NCCGCDB). As a single but high-volume cohort, NCCGCDB included more than 18,000 patients from all around China over the past 20 years. Meanwhile, the US group was identified from the SEER database. The SEER database, supported by the National Cancer Institute, constituted approximately 27.8% of the US population [13]. Younger patients were defined as GC cases younger than 40 years of age, which remains consistent with the majority of previous studies [14,15]. Clinical data abstracted from the NCCGCDB and SEER databases included younger patients' demographics, clinicopathological characteristics, and survival variables. The stage of GC was assessed according to the 8th edition of the American Joint Committee on Cancer (AJCC) TNM staging system. Primary endpoint of the study was overall survival (OS). Moreover, the gene expression sets evaluated in the study were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The study protocol was approved by the ethics committee of National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (No. 17-156/1412), and all methods were carried out in accordance with relevant guidelines and regulations in the Ethics approval and consent.

Statistical Analysis

The line chart was plotted to analyze the changing ratios of younger GC patients from 2000 to 2018. Categorical variables were compared using the chi-squared test. Comparisons were performed using the t-test for normally distributed continuous variables and the Mann-Whitney U test for variables not normally distributed. Survival

curves for different regions and races were calculated with the Kaplan-Meier method, while the log-rank test estimated the relevant survival disparity. Univariate and multivariate Cox proportional hazards models were used to determine the prognostic factors for younger patients, while the corresponding hazard ratio (HR) and 95%CI were generated. The covariates with a *P*-value of < 0.10 in the univariate models were included in the multivariate analysis [16]. Statistical significance was set at a two-sided *P*-value less than 0.05. Survival nomogram for China and the US was formulated based on the multivariate analysis. Younger cases were randomized 7:3, which were adopted in the training set and the validation set, respectively. The area under the curves (AUC), the concordance index (C-index), and the calibration plots were performed to measure the effectiveness of nomograms. All statistical analyses in the study were conducted using R software v.3.6.3 (http://www.r-proje ct.org/).

Bioinformatics Analysis

Based on the GEO database, the differentially expressed genes (DEGs) analysis was further performed to evaluate gene discrepancy for younger patients among different regions, while genes that met the cutoff criteria, adjusted *P*-value＜0.05 and |logFC| ＞ 2.0 were considered as DEGs. Heatmap was constructed with 50 differential genes from DEGs analysis. The Search Tool for the Retrieval of Interacting Genes (STRING) database (http://string-db.org/) was designed to analyze the protein-protein interaction (PPI) network, which was subsequently visualized by Cytoscape software (www.cytoscape.org/). As a plugin in Cytoscape, CytoHubba was used to calculate the degree of protein node, and the top ten genes were identified as hub genes.

RESULTS

Demographic and Clinicopathological features

From 2000 to 2018, a total of 6098 younger GC patients were selected, of which 1159 were enrolled in NCCGCDB, and 4939 were collected from SEER database. The mean age of younger patients was 33.77±4.53 in China, and 32.91±5.38 in the US, respectively (P＜0.0001). As shown in Figure 1, time trends of younger patients from 2000 to 2018

have remained stable in both China and the US sets (range 3.9% - 6.7%, range 3.4% - 3.8%, respectively.). Compare to the US, China group has a higher ratio of younger patients over periods.

Notably, compared to the US, younger GC patients in China were predominant in distal tumor location (83.7% *vs* 66.7%, P<0.0001), poorly differentiation (82.5% *vs* 78.5%, P<0.0001), and signet ring cell carcinoma (52.9% *vs* 28.9%, P < 0.0001). Conversely, relatively higher percentages of proximal location (33.3% *vs* 16.3%, P < 0.0001), well differentiation (7.5% *vs* 1.2%, P < 0.0001), and pTNM stage IV tumors (52.5% *vs* 21.4%, P < 0.0001) were revealed in the US patients (As shown in Table 1). As for treatment, the percentage of surgery (75.1% *vs* 41.4%, P < 0.0001), lymphadenectomy (67.4% *vs* 19.1%, P < 0.0001), and lymphadenectomy with at least 15 lymph nodes (ELNs ≥ 15) (52.4% *vs* 10.6%, P < 0.0001) in China were higher than those in the US. In addition, nearly 60.6% of the patients in the US cohort received chemotherapy, compared with only 36.9% in the Chinese cohort (P < 0.0001).

Survival analysis in different regions and races

Survival trends of younger GC patients were revealed in Table 2 and Figure 2. Compared to the stable prognosis of the US, a noticeable survival increment was shown in younger Chinese patients. The 3-year OS has increased from 54.5% (95%CI: 41.7%-71.4%) in 2000-2003 to 66.5% (95%CI: 60.8%-72.7%) in 2014-2018, while the 5-year OS has improved from 47.7% (95%CI: 35.0%-65.0%) in 2000-2003 to 51.6% (95%CI: 46.6%-57.2%) in 2009-2013.

Figure 3 revealed the Kaplan-Meier curves for different regions and races. The survival outcomes were much better in Chinese or patients diagnosed in China, compared with other races (mentioned in SEER dataset) or the US patients (all P < 0.0001). To avoid the bias of pathological stage, further analysis was performed to evaluate prognosis of younger cases divided by pTNM stage. Left column of Figure 3 showed the survival curves of younger patients between China and the US diagnosed as all pTNM stages (Figure 3A), pTNM stage I (Figure 3B), II (Figure 3C), III (Figure

3D), IV (Figure 3E), while the right column evaluated the OS among different races/ethnicities. Except for stage II GC patients (P = 0.16, P = 0.22, respectively), all other KM curves revealed obvious survival advantages in Chinese or diagnosed in China (all $P < 0.01$).

Moreover, the univariate and multivariate analyses were performed to investigate the prognostic factors for younger GC patients between the China group and the US group (As shown in Table 3 and Table 4).

Significant variables on univariate analysis were enrolled in the multivariate modeling, followed by region, race, period of diagnosis, gender, location, differentiation, linitis plastica, signet ring cell carcinoma, pTNM, surgery, chemotherapy, ELNs ≥ 15 (all $P < 0.05$). For younger GC patients, the independent prognosis factors included region, period of diagnosis, poorly differentiated, linitis plastica, pTNM stage, surgery, chemotherapy, and ELNs≥15 (all $P < 0.05$). However, significant predictors in China only involved period of diagnosis (all $P < 0.05$), linitis plastica (P = 0.0005), and pTNM stage (all $P < 0.0001$). For the US patients, Black ethnicity, poorly differentiated, linitis plastica, signet ring cell, and later pathological stage were related to serious prognosis (all $P < 0.01$), whereas recent period, female, distal location, surgery, and chemotherapy have emerged as protective factors (all $P < 0.05$).

Nomogram of younger GC patients

To predict OS of younger patients with GC, nomograms for China and the US group were established separately based on the results of cox regression analysis. As illustrated in Figure 4, a total of 4 clinical parameters (including linitis plastica, surgery, ELNs ≥ 15, and pTNM stage) were included in the Chinese models, while nomogram for the US identified the following parameters: race, gender, tumor location, differentiation, pTNM stage, linitis plastica, signet ring cell, surgery, and chemotherapy. According to the results of the validation set, the C-index for China and the US models was 0.814 and 0.787. The AUC that was applied to evaluate the discernment of the nomogram models was 0.786 in China group and 0.842 in the US group (Figure S1). In

addition, high-quality calibration plots in both China and the US models had been demonstrated (Figure S2).

DEGs analysis, PPI network, and hub gene identification from GEO database

To evaluate gene disparities according to different races, we performed DEGs analysis for younger GC patients from GEO database. Finally, three gene expression profiles (GSE27342, GSE51105, and GSE38749), which contained 8 younger cases from China, and 12 samples from the Americas and Oceania, were included in this study. After DEGs analysis was performed among different regional sets, significant 50 differently expressed genes were selected and shown in Figure 5. Compared to other regions, certain genes were down-regulated in China (As shown in Figure 5A), while *C11orf58, TLK1, ARHGAP5, MRPL2,* and *RBMX2* were up-regulated. Moreover, the protein interactions among the DEGs were performed with STRING tools. A total of 200 nodes and 245 edges were involved in the PPI network, as presented in Figure 5B and Figure 5C. The top ten genes evaluated by connectivity degree in the PPI network were as follows: *CTNNB1, APP, CTSB, UBA52, RHOA, ATP5B, EEF2, FN1, RPL11,* and *CTSD.*

DISCUSSION

Our study compared younger GC patients between China and the US at clinical and biological levels. Survival advantages of younger GC patients in China were demonstrated from two high-volume databases. Stronger than previous studies [17], high-quality nomograms were further constructed for younger patients according to different regions. Compare to western regions, differential expressed genes in China were confirmed, which might partly explain the histopathological behaviors and survival disparity.

Stayed in line with the previous results [11,18], our conclusions showed that younger patients in China were more prominent in distal location, while GC in proximal location was more common in the US. The above discrepancy was presumably related to certain predisposing factors [19,20,21]. In Western countries, obesity and gastroesophageal reflux

were associated with proximal GC [20], whereas *Helicobacter pylori* infection in China may partially account for a high incidence of distal location [21]. Additionally, Chinese cases were more likely diagnosed with early pathological stage than those in the US. This might partially be owed to the improvement of cancer screening and early detection programs in China, which have expanded to 31 provinces Until 2015 [22,23]. When considering surgical patterns, it was thought that gastrectomy with D2 Lymphadenectomy was the standard treatment for GC in Eastern Asia [24]. However, most patients in Western countries undergo D1 lymphadenectomy, and D2 Lymphadenectomy was only recommended rather than therapeutic norms [25,26]. These treatment differences might explain the higher percentage of ELNs ≥ 15 and the larger number of nodes examined for Chinese younger patients in our study.

With a direct comparison from a high-volume GC cohort in China and the SEER database, we reported significant survival differences for younger GC patients. Consistent with the previous study [11, 18], our findings revealed that younger cases had a better prognosis in China than in the US. In addition to pTNM stage II, significantly better OS was also observed in China group with stages I, II, and IV. Notably, survival advantage of younger Chinese patients could partially be attributed to a remarkable improvement in the quality of clinical services, such as improved access to primary healthcare, early cancer screening, as well as advances in individually multimodal therapies [22, 23, 27]. Moreover, it was well known that gastrectomy with D2 Lymphadenectomy was common in Asian areas, while the vast majority of GC cases in Western countries undergo D1 lymphadenectomy [28]. The above type of surgical procedures might be associated with the survival disparity among different regions. In addition, the later tumor stage, and some other factors including life-habit and high body mass index (BMI) of the US patients might also in part explain the poor prognosis [29].

As a convenient statistical predictive tool, nomogram has been widely applied for physicians to clarify a diagnosis and predict survival [30, 31]. To our best knowledge, the present study, which used the largest sample sizes, constructed survival nomograms for

younger GC patients in China and the US. Compared with previous models [18], the nomograms not only avoid the bias of regions but also achieve reliable predictive performance, which was reflected by an AUC of 0.786 for China and 0.842 for the US. In addition, the calibration plots of the training set and validation set (As shown in Figure S1) illustrated great agreement between nomogram prediction and actual observation, thus further suggesting a robust predictive ability of nomograms in the present study.

Although specific gene expression among regions and races was not yet fully ascertained, several studies demonstrated a strong association between regional/racial-related genes and the prognosis of GC [32, 33]. Loh *et al* [32] review the genetic polymorphisms of GC among different races/ethnicities. The results showed that 37 polymorphisms across 27 genes were significantly related to GC in Asians, while 12 polymorphisms across 11 genes were found in Caucasians. Then, Li *et al* [33] found significant gene disparity among White, Black, and Asian patients. Four core genes, including *GYG2P1, RPS4Y1, TXLNG,* and *EIF1AX,* were demonstrated in White ethnicity, which were relevant for RNA binding and transcription pathways. Black ethnicity with GC was mainly enriched in cell structural changes, and *DNAJC5, HDAC10, NEO1,* and *SMG5* were identified. For Asian patients, *TMSB4Y, UTY, ZFY,* and *ZNF787* were screened based on the relationship between gene expression and DNA methylation. In our study, we first evaluated the race/ethnicity-associated gene for younger GC patients. After the construction of PPI networks and screening according to the degree of the protein node, the hub genes of younger GC patients were as follows: *CTNNB1, APP, CTSB, UBA52, RHOA, ATP5B, EEF2, FN1, RPL11, CTSD.* Further mechanistic studies were warranted to confirm the related signaling and action mechanisms.

Our study has certain strengths. First, two high-volume databases based on the SEER and the NCCGCDB were comprehensively performed to evaluate the regional and racial disparity for younger GC patients. Second, our study, utilizing the largest sample sizes globally, constructed prognostic nomograms for younger GC patients in China and the US, thus providing an intuitive accessible tool for physicians to predict survival

and develop resurvey schedules. Moreover, to our best knowledge, it was the first study to perform biological analysis for younger patients among different regions, which provided a molecular interpretation for the survival discrepancy. Despite all this, several limitations need to be considered in the present study. Some of the key baseline prognostic factors, including BMI, smoking, drinking, and the efficacy of neo/adjuvant therapy were absent in SEER database. In addition, as a high-volume single center, parts of findings from NCCGCDB might not be strongly generalizable to China, which should be replicated in a larger multicenter experience. Lastly, even after correcting by batch effects (As shown in Figure S3), there was also potential bias during encoding samples from GEO database. These factors might affect the accuracy of the results.

CONCLUSION

In conclusion, younger GC patients diagnosed in China had a better prognosis than those in the US. Except for younger cases with pTNM stage II, survival advantage was observed in China group with pathological stages I, II, and IV. Moreover, utilizing the largest sample sizes globally, prognostic nomograms in China and the US were constructed and showed robust predictive performance. Further large-scale studies are warranted to investigate more molecular characteristics and related mechanisms for younger GC patients among different regions and races.

ARTICLE HIGHLIGHTS

Research background

The impact of racial and regional disparity on younger patients with gastric cancer (GC) remained unclear.

Research motivation

This study aimed to provide a national view of younger gastric cancer patients in China and the US.

Research objectives

To investigate the clinicopathological characteristics, prognostic nomogram, and biological analysis of younger gastric cancer patients between China and the US.

Research methods

From 2000 to 2018, GC patients aged less than 40 years were selected from the China National Cancer Center and the Surveillance Epidemiology and End Results (SEER) database. Biological analysis was enrolled from the Gene Expression Omnibus (GEO) database.

Research results

Finally, a total of 6098 younger GC patients were selected from 2000 to 2018, of which 1159 were enrolled in China National Cancer Center, and 4939 were collected from SEER database. Compare with the US group, younger patients in China revealed better survival outcomes ($P < 0.01$). For race/ethnicity, younger Chinese cases also enjoyed a better prognosis than that in White and Black sets ($P < 0.01$). Prognostic nomograms for younger patients were established, with an AUC of 0.786 for China and 0.842 for the US. Moreover, three gene expression profiles (GSE27342, GSE51105, and GSE38749) were enrolled in further biological analysis, and distinctive molecular characteristics were identified in younger GC patients in different regions.

Research conclusions

Except for younger cases with pTNM stage II, survival advantage was observed in China group with pathological stages I, II, and IV. Biological analysis of younger patients was performed among different regions, which might partly explain the histopathological behaviors and survival disparity in the subpopulations.

Research perspectives

Further large-scale studies are warranted to investigate more molecular characteristics and related mechanisms for younger GC patients among different regions and races.

ACKNOWLEDGEMENTS

We thank Zhou Dawei (Qilu Hospital, Cheeloo College of Medicine, Shandong University) for assistance with data analysis.

79487_Auto_Edited.docx

ORIGINALITY REPORT

4%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.researchsquare.com Internet	46 words — 1%
2	www.ncbi.nlm.nih.gov Internet	29 words — 1%
3	www.dovepress.com Internet	25 words — 1%
4	Hua Lin, Hongwei Liu. "Hypoxia on vhl-deficient cells to obtain hif related genes through bioinformatics analysis", Cold Spring Harbor Laboratory, 2019 Crossref Posted Content	14 words — < 1%
5	link.springer.com Internet	14 words — < 1%
6	Lulu Zhao, Penghui Niu, Dongbing Zhao, Yingtai Chen. "Regional and racial disparity in proximal gastric cancer survival outcomes 1996–2016: Results from SEER and China National Cancer Center database", Cancer Medicine, 2021 Crossref	13 words — < 1%
7	e-century.us Internet	13 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 12 WORDS
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 12 WORDS