



OPEN

A comparison of Chinese multicenter breast cancer database and SEER database

Yingjie Niu^{1,2,6}, Fan Zhang^{3,6}, Dong Chen^{4,6}, Guolin Ye^{5,6}, Yong Li¹, Yong Zha¹, Wenlin Chen², Dequan Liu², Xiaoming Liao³, Qinghua Huang³, Wei Tang³, Gengxi Cai⁵, Rong Guo²✉, Hongyang Li¹✉ & Shicong Tang²✉

There are different characteristics of BC in developing countries and developed countries. We intended to study the factors which influence the survival and prognosis of BC between southern China and the United States. (a) To study the two groups BC patients in southern China from 2001 to 2016 and SEER database from 1975 to 2016. (b) To register, collect and analyze the clinicopathological features and treatment information. Our study found that there are significant differences in tumor size, positive lymph node status and KI-67 between southern China and SEER cohort ($P < 0.000$). The positive lymph node status may be one of the causes of difference of morbidity and mortality of BC patients in China. Furthermore, the differences in treatment methods may also account for the differences between China and seer databases.

Breast cancer (BC) is one of the most common cancers of women in the world, which is also the second main cause of cancer death in women. There are about 1.5 million women (25% of all cancer women) are diagnosed with BC every year all around the world, and it is estimated to increase to 2.2 million per year by 2025¹⁻³. However, there are literatures showed that the mortality of BC in high-income countries is declining, while it is increasing in low and middle income countries⁴.

It is reported that, the incidence of BC in China is expanding at twice the rate of the world, which has risen from 23.37/10 million in 2007 to 28.42/10 million in 2013^{5,6}. In addition, there are researches showed that the incidence rate of BC is increasing in China, and it may reach 2 million 500 thousand cases by 2021, however, its survival rate is low, the 5-year survival rate ranged from 40 to 60%^{7,8}. Then, a lot of research suggests that there is a large difference of BC between China and other regions, which may due to the difference of clinicopathological features, the details as follows⁹⁻¹¹. Jarzab et al. considered that the prognosis of lumen type G1 tumor is good, while G3 tumor is poor in BC of China¹². Meanwhile, Zhang et al. reported that, lymph node metastasis seriously affects the prognosis of BC patients. In addition, the expression of KI-67 and prognosis were closely related to pTNM stage and PR expression¹³. It is indicated that Ki-67 positive can lead to higher histological grade of BC¹⁴. However, HER2 overexpression can lead to invasive breast cancer, which overexpression was negatively correlated with the expression of PR and ER¹⁵.

In contrast, there are about 12% of women in the United States are diagnosed with BC in their lifetime, nevertheless, it is estimated that there are about 3.1 million BC survivors each year^{16,17}. With the development of treatment strategies, the mortality of BC has been decreased in the United States, which the 5-year survival rate was about 90% after treatment¹⁸. A large number of studies have shown that, BC also has its own characteristics in the United States. DeSantis CE et al. reported that BC patients in the United States from 2004 to 2014, young women have higher invasive and specific genomic characteristics, meanwhile, the incidence of HR positive (ER positive or PR positive) BC increased, while the morbidity of HR negative tumors decreased¹⁹⁻²¹.

¹Department of Hepatobiliary and Pancreatic Surgery, The Third Affiliated Hospital of Kunming Medical University, Yunnan Cancer Hospital, Kunming, Yunnan, People's Republic of China. ²Department of Breast Surgery, The Third Affiliated Hospital of Kunming Medical University, Yunnan Cancer Hospital, Kunming, Yunnan, People's Republic of China. ³Department of Breast Surgery, Affiliated Tumor Hospital of Guangxi Medical University, Nanning, Guangxi, People's Republic of China. ⁴Department of Ultrasound, Cancer Hospital of Yunnan Province, The Third Affiliated Hospital of Kunming Medical University, Kunming 650118, Yunnan, People's Republic of China. ⁵Department of Breast Surgery, The First People's Hospital of Foshan, #81, North Lingnan Avenue, Chancheng, Foshan, Guangdong, People's Republic of China. ⁶These authors contributed equally: Yingjie Niu, Fan Zhang, Dong Chen and Guolin Ye. ✉email: guorong2320@126.com; 1564356100@qq.com; tang_shicong@126.com

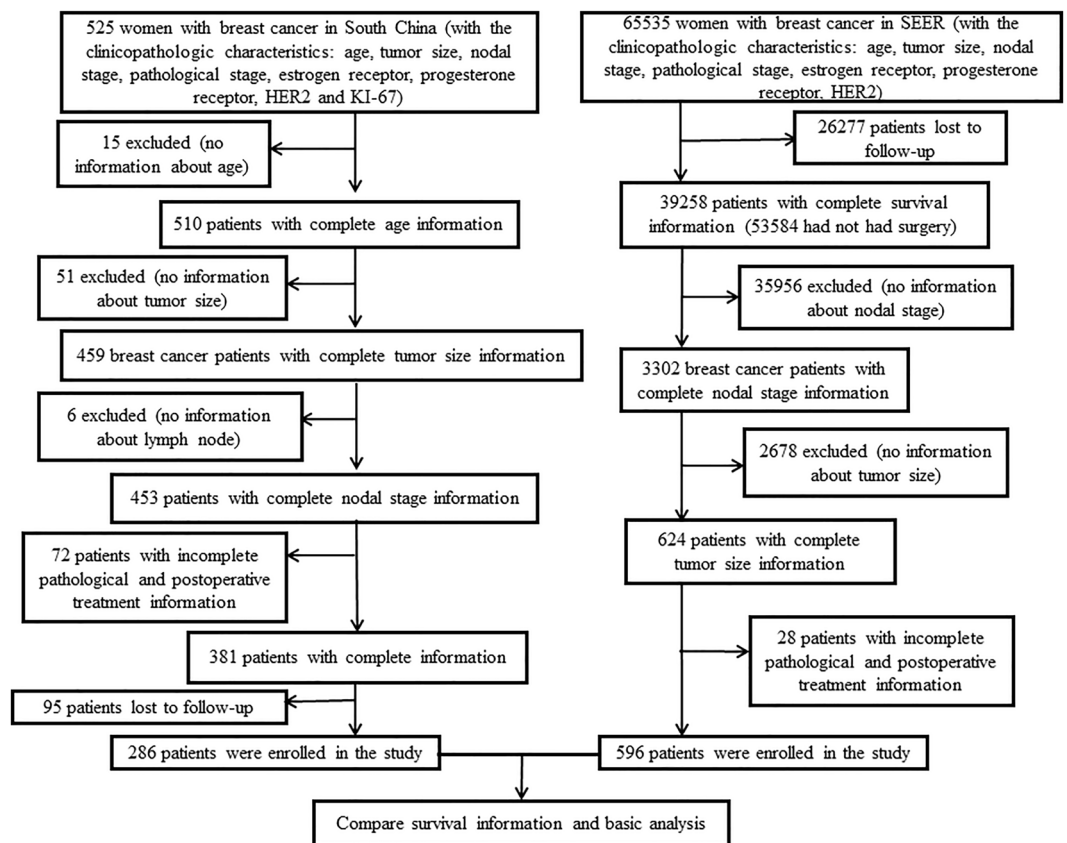


Figure 1. Flow chart for breast cancer patients in both Southern China and SEER database.

There are different researchers in the world comparing BC in China and other regions, and finding that there exist some differences between them. However, the comparison of BC patients between southern China and the United States has not been reported. This study aims to investigate the differences of BC patients between China and the population-based Surveillance, Epidemiology, and End Results (SEER) cohort. In addition, our study intends to examine the age, stage and grade of tumor, ER, PR, HER2, KI-67 and treatment methods, in order to analyze the age distribution, clinical characteristics, treatment and prognosis of BC patients in Chinese multicenter breast cancer database and SEER database, so that we could compare the two groups.

Methods

Patients and ethics. We conducted a retrospective analysis and comparison of the patients who have been diagnosed with primary breast cancer in southern China (2001–2016) and SEER database (1975–2016). Overall, there was a total of 525 breast cancer patients were diagnosed in southern China, among them, 15 patients were excluded from this study due to lack of age information. In addition, 129 patients were removed, which without tumor stage, ER, PR, HER2, KI-67 and treatment information. Additionally, there are about 95 patients were lost. Finally, a total of 286 patients were included in the study (Fig. 1). The study was approved by Institutional Review Board of Yunnan Cancer Hospital, Cancer Hospital Affiliated to Guangxi Medical University and Foshan first people's Hospital. Informed consent was obtained from all individual participants included in the study. All procedures implemented in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee, and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The SEER cohort was derived from the SEER database (November 2018 submission) by using SEER*Stat software provided by the National Cancer Institute (NCI). There were 65,535 breast cancer patients, among them, 26,277 cases were lost. Except for without complete information (age, stage and grade of tumor, ER, PR, HER2, KI-67 and treatment) patients, which was about 38,662 cases, there were 596 patients included in this study (Fig. 1).

Clinical data collection. A retrospective review of medical records and pathology reports was conducted. Staging was performed according to the American Joint Committee on Cancer (AJCC) guidelines²². The age of the patients was classified to young adult group (<40), middle aged group (40–70) and aged + group (>70), and then, calculated the median age of each age group (35, 48, 75), finally, statistical analysis was carried out respectively. A cutoff of 14% for KI-67 was used, which was recommended by 2011 St Gallen consensus panel²³, and then, we divided the group of KI-67 $\geq 14\%$ into two subgroups, according to the median (51.7% in southern China). Patients in southern China were told to have an examination and treatment according to the guidelines

of the breast cancer center and were followed up by telephone, and collect the information about survival and treatment, including date of progression metastasis, date of relapse and date cause of death.

Statistical methods. The IBM SPSS Statistics (Version 21.0; IBM Corp., New York, USA) and GraphPad Prism (Version 6.0; GraphPad software, Inc., LaJolla, CA, USA) were used for statistical analysis. Disease-free survival (DFS) was measured from the beginning of the operation to the first recurrence/metastasis of the tumor or the death of the subject for any reason (the last follow-up time was the patients who lost the follow-up; the patients who were still alive at the end of the study were the end of the follow-up). Overall Survival (OS) was measured from the beginning of operation to death due to any reason. Univariate analysis and multivariate analysis was performed by Cox regression analysis, according to comparing the age (<40 and \geq 40), tumor size (\leq 2 cm and >2 cm), node status, ER, PR, HER2, KI-67, surgery and radiation. Kaplan–Meier method was used to estimate DFS and OS, log-rank test was used to compare the patients with different clinicopathologic characteristics. The count data were tested by χ^2 test, fisher exact probability method was used when the cases was less than 6. Statistical significance was set at a $P < 0.05$, $P < 0.01$ had significant difference.

Ethics approval and consent to participate. All procedures implemented in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee, and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Ethics/Institutional Review Board approval of research—Yunnan Cancer Hospital, Cancer Hospital Affiliated to Guangxi Medical University and Foshan first people's Hospital. Informed consent was obtained from all individual participants included in the study.

Consent for publication. All patients enrolled in the study signed the consent for publication.

Results

Patients. Between 2001 and 2016, a total of 525 females received surgical treatment in the southern China. Among them, 15 patients without the age information, so they were excluded from this study. There were 51 patients without the information of tumor size, 6 patients without node status, and in the left 453 cases, 75 patients without pathological stage and treatment information and 95 patients were lost, all in all, there were 286 patients included in the southern China cohort. Between 1975 and 2016, a total of 65,535 patients were included in SEER database, among them, 26,277 patients were lost to follow-up, 35,956 patients without the information of lymph node stage, 2678 cases had no tumor size information, 28 patients had no complete information of tumor staging and treatment, so there were about 596 patients included from the SEER cohort (Fig. 1).

Comparison of clinicopathological features between southern China and SEER cohort. All patients were divided into three subgroups: young adult group (<40), middle aged group (40–70) and aged + group (>70). Among 510 breast cancer patients in southern China, 101 (19.8%) patients were under 40, what's more, middle aged patients account for the most, about 396 (77.65%) patients. Between 1975 and 2016, there were 65,535 BC patients were included SEER database, the proportion of young patients was slightly less than that in southern China, was 4024 (6.14%), but there was no statistical significance between the two ($P = 0.923$). However, the middle aged patients account for 41,256 (62.95%) in SEER cohort ($P = 0.000$). There was a difference between southern China and SEER cohort, the proportion of aged + group was higher in SEER cohort, which was about 13,104 (20.00%), it was significantly higher than 5 patients (0.98%) in southern China, which was statistically significant ($P = 0.048$) (Tab. 1). BC patients in southern China and SEER cohort were compared, the age of newly diagnosed cases in the two groups was 40–70 years old. While, there was a different between the two groups: BC patients in the subgroup of 40–48 years old (48 years old was the median age of 40–70 years old) were roughly similar to the subgroup of 48–70 years old in southern China, there were 194 (38.04%) patients and 202 (39.61%) patients, respectively. However, BC patients in SEER cohort were slightly older, it was 9105 (13.89%) patients in the subgroup of 40–48 years old, while there were 32,151 (49.06%) patients in the subgroup of 48–70 years old (Table 1).

According to TNM stage, BC patients were divided into Tx, Tis and invasive group (including \leq 2 cm, $2 < 5$ cm, > 5 cm), among them, invasive group account for the most, there were 448 (89.96%) patients in southern China and 1364 (2.99%) patients in SEER cohort, $P = 0.000$. In southern China, the tumor size was 2–5 cm accounted for the most, were 323 (64.86%) patients. However, there were too many data were missing, were 64,071 (97.8%) cases, Tis subgroup in the two was both 1 case (Table 1). There were 257 (50.39%) patients with node metastasis in southern China and 296 (0.45%) cases in SEER cohort. Among the two groups, no lymph node metastasis accounted for the most, were about 206 (40.39%) patients and 1145 (1.75%) patients. While, there were too many data of lymph node status missing, were 29,269 (44.7%) patients (Table 1). Comparing southern China and SEER database, there was statistical significance in tumor stage ($P = 0.000$). Among them, the proportion of stage 2 was the highest, were 262 (51.37%) cases and 14,936 (22.81%) cases respectively. Next, it was stage 3, were 133 (26.08%) cases and 12,993 (19.75%) cases. However, there were many missing data about tumor stage in SEER database, were about 29,269 (44.7%) cases (Table 1). The expression of ER was counted in both southern China and SEER cohort ($P = 0.000$). Among them, ER (+) accounted for a higher proportion, were 290 (56.86%) cases and 20,233 (65.34%) cases respectively, and ER (–) were 178 (34.9%) cases and 5563 (17.96%) cases, respectively (Table 1). Similarly, the differences in expression of PR in southern China and SEER cohort was also statistically significant ($P = 0.000$). Among them, PR (+) was higher in both southern China and SEER cohort, were 270 (52.94%) cases and 179,194 (55.55%) cases, PR (–) were 182 (35.69%) cases and 8110 (26.2%) cases (Table 1). While, the expression of HER2 was different between the two groups ($P = 0.000$). The

Patient characteristics	Southern China	SEER	χ^2 -value	P value
	N (%)	N (%)		
Age at diagnosis (year)			0.009	0.923
Young adult group (<40)				
< 35	38 (7.45)	1495 (2.28)		
≥ 35	63 (12.35)	2529 (3.86)		
Middle aged group (40–70)			163.919	0.000
≤ 48	194 (38.04)	9105 (13.89)		
> 48	202 (39.61)	32,151 (49.06)		
Aged + group(> 70)			0.077	0.048
≤ 75	8 (1.57)	7151 (10.91)		
> 75	5 (0.98)	13,104 (20.00)		
Tumor size(cm)			29.923	0.000
Tx	10 (2.01)	51 (0.08)		
Tis	1 (0.20)	1 (0)		
Invasive group				
≤ 2	79 (15.86)	934 (1.43)		
> 2 ≤ 5	323 (64.86)	367 (0.56)		
> 5	46 (9.24)	63 (0.1)		
Missing	39 (7.83)	64,071 (97.8)		
Nodal stage			20,387.140	0.000
pN0, no nodal metastasis	206 (40.39)	1145 (1.75)		
pN1, 1–3 nodal metastasis	132 (25.88)	222 (0.34)		
pN2, 4–9 nodal metastasis	76 (14.90)	34 (0.05)		
pN3, ≥ 10 nodal metastasis	49 (9.61)	40 (0.06)		
Unknown	47 (9.22)	64,094 (97.80)		
Pathological stage			25.662	0.000
0–1	97 (19.02)	7082 (10.82)		
2	262 (51.37)	14,936 (22.81)		
3	133 (26.08)	12,993 (19.75)		
4	18 (3.53)	1255 (1.92)		
Unknown	0 (0)	29,269 (44.7)		
Estrogen receptor			106.162	0.000
Negative	178 (34.9)	5563 (17.96)		
Positive	290 (56.86)	20,233 (65.34)		
Missing	42 (8.24)	5172 (16.70)		
Progesterone receptor			30.794	0.000
Negative	182 (35.69)	8110 (26.2)		
Positive	270 (52.94)	179,194 (55.55)		
Missing	58 (11.37)	5646 (18.24)		
HER2-positive			611.595	0.000
Negative	165 (32.35)	1530 (85.62)		
Positive	283 (55.49)	169 (9.46)		
Missing	62 (12.16)	88 (4.92)		
Ki-67(%)			NA	NA
< 14	100 (19.61)	NA (NA)		
< 51.7 median	170 (33.33)	NA (NA)		
≥ 51.7 median	95 (18.63)	NA (NA)		
Missing	145(28.43)	NA (NA)		

Table 1. Clinicopathologic characteristics of breast cancer patients in Southern China and SEER cohorts.

expression of HER2 (+) accounted for the most in southern China, was 283 (55.49%) cases, but there was only 169 (9.46%) cases in SEER cohort, with the proportion of PR (–) was high, was 1530 (85.62%) cases (Table 1). Additionally, the expression of KI-67 was different in the two groups, there were 365 (71.57%) cases of KI-67 (+), among them, it was the most between 14 and 51.7%, was 170 (16.58%) cases. However, there was no data about the expression of KI-67 in SEER cohort (Table 1).

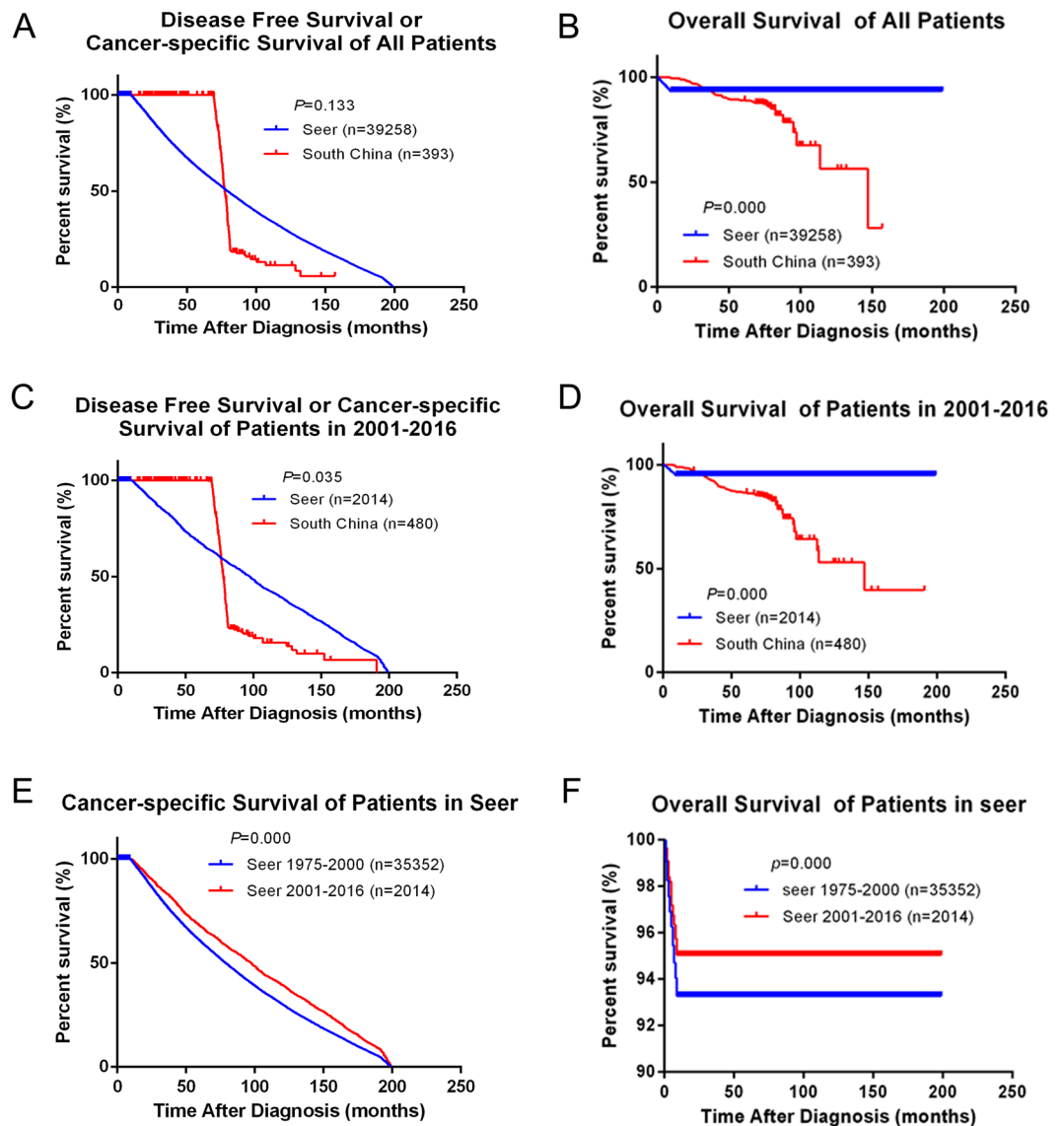


Figure 2. Comparison of all patients' DFS, CSS and OS in Southern China and SEER database. (A,B) Kaplan-Meier estimates disease-free survival (DFS), cancer-specific survival (CSS) and overall survival (OS) of all patients in both Southern China and SEER database, (C,D) Kaplan-Meier estimates DFS or CSS and OS of patients in 2001–2016 in both Southern China and SEER database, (E,F) Kaplan-Meier estimates DFS or CSS and OS in both 1975–2000 and 2001–2016 in SEER database.

Survival and prognosis analysis. Between 2001 and 2016, there were 393 BC patients who were followed up in southern China. Meanwhile, between 1975 and 2016, 39,258 BC patients with complete follow-up information were analyzed. During the follow-up period, there were 59 patients died, 219 patients were alive and 115 patients relapsed in southern China. While in SEER cohort, there were 3613 patients died and 35,645 patients were alive. DFS, cancer-specific survival (CSS) and OS of all included breast cancer patients were compared: in comparing of DFS or CSS, there was no significant difference between southern China and SEER cohort ($P=0.133$), but there was significant difference in OS ($P=0.000$), and the OS in SEER cohort. It was significantly higher than southern China (Fig. 2A,B). Secondly, since the data of southern China was only included from 2001 to 2016, a layered statistic was used to count the DFS or CSS and OS in both southern China and SEER database from 2001 to 2016. Among them, in the first 70 months of follow-up, DFS in southern China was higher than CSS in SEER cohort, and then, CSS in SEER cohort was significantly higher than that in southern China ($P=0.035$), and OS in this period also has significant statistical difference ($P=0.000$), SEER cohort was significantly higher than that in southern China (Fig. 2C,D). Finally, SEER cohort was analyzed in stages, dividing into 1975–2000 and 2001–2016 two subgroups, furthermore, CSS and OS of each subgroup were counted respectively. The results showed that, in SEER cohort, the DFS or CSS and OS of 2001–2016 were significantly higher than 1975–2000 ($P=0.000$) (Fig. 2E,F).

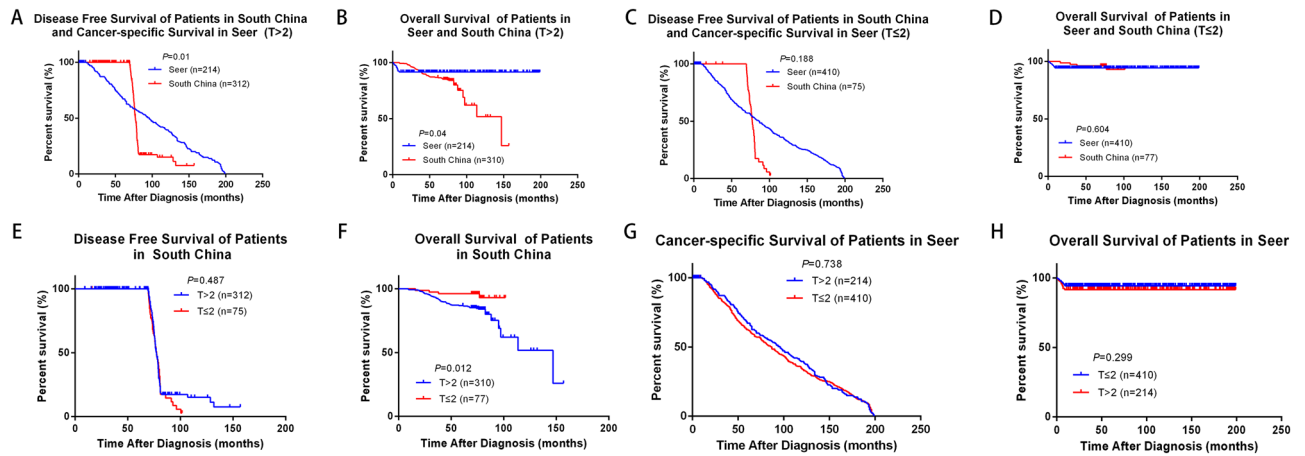


Figure 3. Comparison of different tumor size in Southern China and SEER database. (A,B) Kaplan–Meier estimates DFS or CSS and OS of patients whose T > 2 cm in both Southern China and SEER database, (C,D) Kaplan–Meier estimates DFS or CSS and OS of patients whose T ≤ 2 cm in both Southern China and SEER database, (E,F) Kaplan–Meier estimates DFS and OS of different tumor size in Southern China, (G,H) Kaplan–Meier estimates CSS and OS of different tumor size in SEER database.

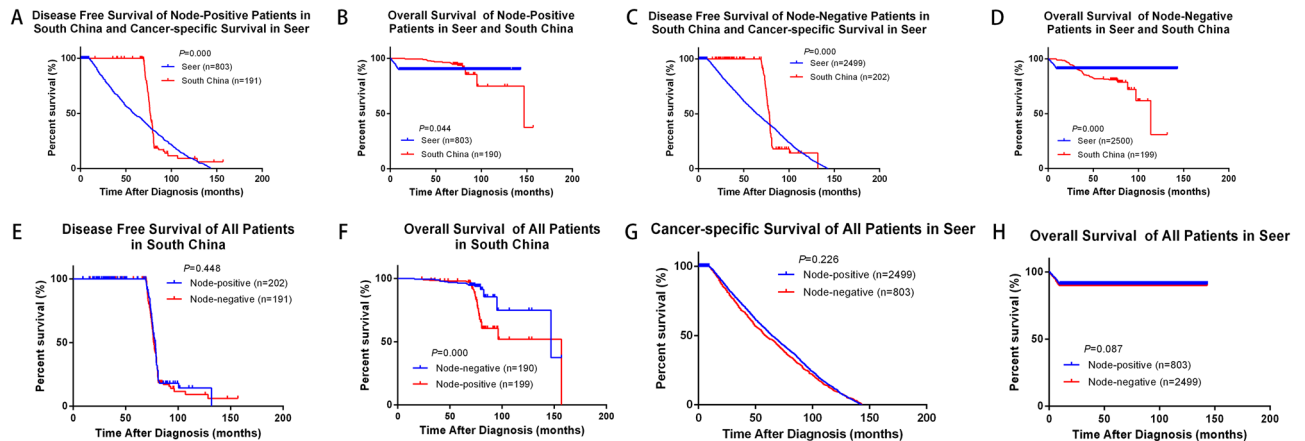


Figure 4. Comparison of different Nodal stage in Southern China and SEER database. (A,B) Kaplan–Meier estimates DFS or CSS and OS of Node-positive patients in both Southern China and SEER database, (C,D) Kaplan–Meier estimates DFS or CSS and OS of Node-negative patients in both Southern China and SEER database, (E,F) Kaplan–Meier estimates DFS and OS of different nodal stage in Southern China, (G,H) Kaplan–Meier estimates CSS and OS of different nodal stage in SEER database.

We analyzed the influence of different clinicopathological features on survival and prognosis of BC patients in southern China. By analyzing the effects of age, tumor size, lymph node status, ER, PR, HER2, KI-67, surgery and radiotherapy on the prognosis of breast cancer, we found that tumor size, positive lymph node status and KI-67 expression affected OS of BC patients in southern China, which showed significant statistical difference ($P = 0.018$, $P = 0.000$, $P = 0.034$ respectively) (Supplementary Fig. 1). We further analyzed and compared the effect of different tumor size on survival of different BC cohorts. There were statistical differences of DFS or CSS and OS in SEER cohort and southern China when the tumor size (T) > 2 cm ($P = 0.01$ and $P = 0.04$), however, DFS or CSS and OS were not statistically different in the two groups when $T \leq 2$ cm ($P = 0.188$ and $P = 0.604$) (Fig. 3A,D). Secondly, the effects of different tumor sizes on the survival of BC patients in each cohort were analyzed separately. Among them, tumor size had little effect on DFS in southern China ($P = 0.487$), but for OS, there was significant statistical difference, OS in T > 2 cm group was significantly lower than T ≤ 2 cm ($P = 0.012$) (Fig. 3E,F). However, for SEER cohort, CSS and OS of T > 2 cm group were slightly lower than that of T ≤ 2 cm group, but there was no statistical difference ($P = 0.738$ and $P = 0.299$) (Fig. 3G,H). We analyze and compared the effect of different node stage on survival of different BC cohorts. Positive-node affected DFS or CSS and OS in both southern China and SEER cohort ($P = 0.000$ and $P = 0.044$). Meanwhile, negative-node also affected DFS or CSS and OS in the two groups ($P = 0.000$ and $P = 0.000$). OS of SEER cohort with different lymph node status was higher than that of southern China (Fig. 4A–D). Analyzing southern China and SEER cohort separately, DFS or CSS and OS of positive-node were lower than negative-node, among them, OS of lymph node status has significant statistical difference ($P = 0.000$), but DFS or CSS of lymph node status has no statistical difference

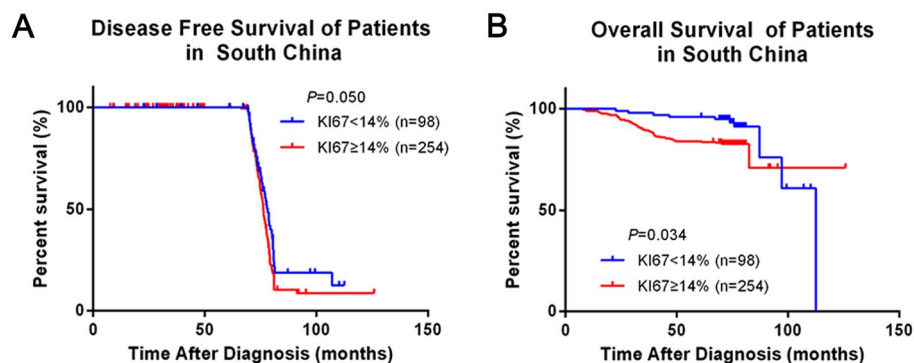


Figure 5. Comparison of different KI-67 expression in Southern China. (A) Kaplan–Meier estimates DFS of different KI-67 expression in Southern China, (B) Kaplan–Meier estimates OS of different KI-67 expression in Southern China.

Variables	Univariate analysis			Multivariate analysis		
	Hazard ratio	95% confidence interval	P value	Hazard ratio	95% confidence interval	P value
Age	0.874	0.644–1.186	0.388	0.977	0.600–1.591	0.926
Tumour stage > 2 cm	0.894	0.652–1.227	0.488	1.257	0.793–1.990	0.330
Node-positive	1.108	0.850–1.446	0.448	1.116	0.754–1.653	0.584
Estrogen receptor-positive	1.109	0.846–1.454	0.453	1.154	0.660–2.017	0.615
Progesterone receptor-positive	1.116	0.852–1.463	0.426	1.122	0.662–1.904	0.669
HER2-positive	0.885	0.675–1.162	0.380	1.091	0.740–1.608	0.662
Ki-67 high expression	1.376	1.000–1.894	0.050	0.960	0.385–2.392	0.930
Mastectomy	0.934	0.521–1.673	0.818	0.985	0.650–1.492	0.943
Radiation therapy	1.016	0.770–1.342	0.910	0.889	0.596–1.326	0.564

Table 2. Cox regression analysis for disease-free survival in southern China.

($P=0.448$) (Fig. 4E,F). But for SEER cohort, CSS and OS of positive-node was slightly higher than negative-node, while there was no statistical difference ($P=0.226$ and $P=0.087$) (Fig. 4G,H). We analyzed and compared the effect of expression of KI-67 on survival of southern China. Among the subjects included in this study, DFS and OS of KI-67 < 14% both higher than $\geq 14\%$, there were significant statistical difference ($P=0.05$ and $P=0.034$) (Fig. 5A,B). Multivariate analysis and univariate analysis of southern China and SEER cohort was performed by Cox regression analysis. In univariate analysis of DFS, T > 2 cm, positive-node, ER (+), PR (+), HER2 (+), surgery and radiation all had no significant influence on the increased risk of death. Among them, the hazard ratio (HR) of KI-67 high expression group was 1.376, 95% CI 1.000–1.894, $P=0.050$. However, in multivariate analysis of DFS, all the clinicopathological features of the included studies were statistically significant (Table 2).

Multivariate analysis and univariate analysis of southern China was performed by Cox regression analysis. In the univariate analysis of OS, it was significantly associated with increased risk of death and T > 2 cm (HR 3.406, 95% CI 1.232–9.417, $P=0.018$), positive-node status (HR 0.308, 95% CI 0.169–0.564, $P=0.000$) and KI-67 high expression (HR 2.128, 95% CI 1.057–4.285, $P=0.034$). In the multivariate analysis of OS, positive-node status (HR 0.226, 95% CI 0.098–0.519, $P=0.000$) was significantly associated with increased risk of disease survival (Table 3).

Comparison of treatment. There were differences of treatments between southern China and SEER cohort (Table 4). We analysed both of the two databases from 2001 to 2016, and there were 389 (97.01%) patients received chemotherapy in southern China, but there were 1574 (26.64%) patients in SEER cohort have received chemotherapy ($P=0.000$). However, the treatment of surgery was similar, there were 387 (95.09%) patients performed mastectomy in southern China cohort (including simple resection and modified radical operation), and for SEER cohort, there were 5524 (93.50%) patients had breast surgery (the specific operation method is not clear). However, it was the same of the two about whether to receive radiotherapy or not, and there was a significant statistical difference ($P<0.001$). Among them, there were 351 (72.82%) patients had not performed radiotherapy in southern China, with 2223 (37.63%) patients in SEER cohort.

Changes in morbidity with years. We further analyzed and compared the age distribution of BC patients in different years in SEER cohort. Among them, except for 90's, the proportion of BC patients in young adult group (<40 years old) showed an increasing trend with age. They were respectively: 70's: 6%, 80's: 6.31%, 90's:

Variables	Univariate analysis			Multivariate analysis		
	Hazard ratio	95% confidence interval	P value	Hazard ratio	95% confidence interval	P value
Age	0.771	0.462–1.288	0.321	0.836	0.386–1.814	0.651
Tumour stage > 2 cm	3.406	1.232–9.417	0.018	0.710	0.246–2.049	0.526
Node-positive	0.308	0.169–0.564	0.000	0.226	0.098–0.519	0.000
Estrogen receptor-positive	0.716	0.454–1.127	0.149	1.888	0.702–5.079	0.208
Progesterone receptor-positive	0.647	0.404–1.035	0.069	0.561	0.207–1.518	0.255
HER2-positive	0.945	0.599–1.490	0.808	1.576	0.825–3.009	0.168
Ki-67 high expression	2.128	1.057–4.285	0.034	1.013	0.115–8.902	0.990
Mastectomy	3.351	0.463–24.256	0.231	1.020	0.501–2.077	0.957
Radiation therapy	1.088	0.677–1.748	0.727	1.130	0.584–2.188	0.717

Table 3. Cox regression analysis for overall survival in southern China.

Treatment	Southern China	SEER	χ^2 -value	P value
	N (%)	N (%)		
Chemotherapy			867.474	< 0.001
Yes	389 (97.01)	1574 (26.64)		
No	12 (2.99)	4334 (73.36)		
Breast surgery			NA	NA
Breast-conserving surgery	20 (4.91)	5524 (93.50)		
Mastectomy	387 (95.09)			
No breast surgery	0 (0)	384 (6.50)		
Radiation therapy			229.469	< 0.001
Yes	131 (27.18)	3685 (62.37)		
No	351 (72.82)	2223 (37.63)		

Table 4. Treatment of breast cancer patients in Southern China and SEER cohorts.

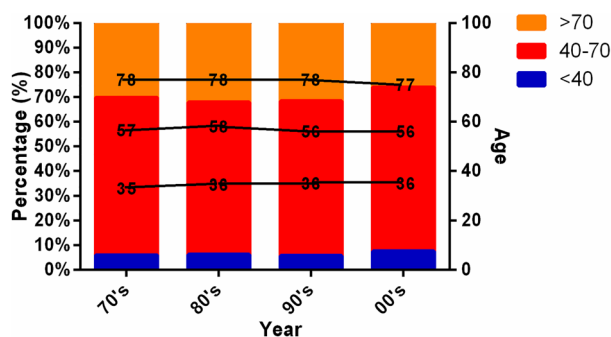


Figure 6. Comparison of different years all patients' age in SEER database.

5.76%, 00's: 7.62%, and the median age was 35 year, 36 year, 36 year and 36 year. In all age groups, the incidence of middle-aged group (40–70 years old) was the highest, followed by 63.89%, 61.74%, 62.67% and 66.42%, and the median age was 57 year, 56 year, 56 year and 56 year. In addition, we found that the morbidity of aged + group (> 70) showed a decreasing trend year by year, respectively was 30.11%, 31.95%, 31.57% and 25.96%, and the median age was orderly 78 year, 78 year, 78 year and 77 year (Fig. 6).

Discussion

Our research is a very important one, which is the first to analyze and compare the related factors of survival and prognosis of BC patients in southern China and SEER cohort. We analyzed the multiple factors: including age, tumor stage and grade, ER, PR, HER2, KI-67, surgery and radiotherapy, which may influence the survival and prognosis of BC. It is a multi-regional and big data clinical study.

In this study, by comparing and analyzing the age of both southern China and SEER cohort, we found that in southern China from 2001 to 2016, there was about 19.8% BC patients were under 40 year, which was the same as the results of Wang, who had reported that the incidence of young BC patients in China is about 21.97%²⁴. However, the morbidity of young BC patients in China is significantly higher than that in western countries (about 4–6%)^{25–28}, which was similar to the incidence of our study: it was 6.14% of SEER cohort from 1975 to 2016. All of these suggested that the incidence of BC in China is younger than that in western countries, which indicated that age may be a factor affecting the survival and prognosis of BC patients in southern China and the United States.

To further study DFS and OS, we focused on T stage, positive lymph node status, ER, PR, HER2, KI-67 expression of BC patients, and thought that T stage, positive lymph node status and KI-67 expression all could be regarded as factors, which affected the survival and prognosis of BC patients. Other scholars had also studied tumor stage, and found that there were about 60–70% of BC patients were diagnosed with stage 1, which was higher than Asian countries, but there were only about 10% women were stage 4²⁹. This research was similar to our results, our finding showed that most BC patients in southern China from 2001 to 2016 could be diagnosed at early time, among them, there were about 64.86% of patients with T2 BC, 40.39% of patients with N0, 51.37% of patients with stage 2. However, the early diagnosis rate of BC in China is far lower than that in the United States. It showed that in SEER cohort, there were about 65.96% of patients with T1, 79.46% of patients with N0 and 41.18% of patients with stage 2. Cox regression analysis showed that T stage and positive lymph node status were important factors affecting OS of BC. Therefore, we further to prove that stages and grades of tumor had a significant impact on the survival and prognosis of BC. Meanwhile, China should to further strengthen the early diagnosis and treatment of BC, so as to improve the prognosis of BC patients in China.

We further explored the effects of ER, PR, HER2, KI-67 expression on survival and prognosis of BC. The proportion of ER (+) BC patients was similar in both southern China and SEER cohort. It was 56.86% in southern China and 65.34% in SEER cohort, which was slightly lower than that had been reported (about 70%)³⁰. It may be related to excessive data deletion in SEER cohort, which was about 16.7% of ER data were missing in this study. Other studies have showed that the most important factors affecting the prognosis of BC were tumor grade and ER status³¹. However, in our study, ER was not an indicator of survival and prognosis of BC. Additionally, there were different treatment methods of BC according to the different status of hormone receptor (HR). Endocrine therapy could be used for ER or PR positive patients, but the effect of chemotherapy was not as good as these of negative patients, and the different treatment methods could significantly affect the prognosis of BC. PR was also an important factor affecting the prognosis of BC. In our study, there was 52.94% of PR positive in southern China, and 55.55% of that in SEER cohort, there was significant statistical difference between the two groups ($P=0.000$). Similarly, the results of Cox regression analysis showed that PR was also not an indicator of survival and prognosis of BC, which may had a relationship between a large data missing. In addition, Ding et al. found that BC with HER2 and KI-67 overexpression had higher lymph node metastasis rate and higher AJCC tumor stage^{32,33}, which was similar to our results. In this study, HER2 positive were 55.49% and 9.46% in southern China and SEER cohort respectively ($P=0.000$), which was consistent with the literature showing that BC cells from young patients are more likely to show HER2 positive expression²⁴. KI-67 positivity was 70.01% in southern China, a high expression is an important factor affecting OS in BC patients. However, DFS was detected by χ^2 test when KI-67 was regarded as an independent factor, $P=0.05$, but we believe that this was mainly due to the small sample size, the trend in the conclusion was still valid, as the sample size continues to increase, the value of P may gradually decrease. In summary, our results showed that positive-node status was an important factor affecting the prognosis of BC, which also reflected that BC patients in southern China and the United States have different biological behaviors and pathogenesis.

Additionally, the treatment methods of BC were also important factors affecting its prognosis. At present, the main treatments of BC were surgery, radiotherapy, chemotherapy, targeted therapy and hormone therapy^{34,35}. Among them, surgery can significantly reduce the mortality rate, which is the most critical step in the treatment of breast cancer, there are five common surgical methods: breast conserving surgery (BCS), simple mastectomy (SM), modified radical mastectomy (MRM), radical mastectomy (RM) and extensive radical mastectomy (ERM)³⁶. Among them, Bartelink et al. reported that BCS has the equivalence with mastectomy³⁷. However, the comparison of treatment methods between southern China and the United States has not yet reported, we had studied this for the first time. In this study, BC patients enrolled in the study in southern China all underwent surgery, and the treatment including BCS and mastectomy (including SM and MRM). Among them, there were about 95.09% of BC patients performed mastectomy and 4.91% of that performed BCS, and the implementation rate of BCS was significantly lower than that of developed countries, which was similar to the results of Gupta A: the highest implementation rate of BCS in China is only 8.6%³⁸. However, there were only 18.24% of BC patients performed mastectomy in SEER cohort, 81.76% of that had not. While, there was no significant effects on DFS and OS for whether performing surgery, which suggesting that surgery had little effect on the survival and prognosis of BC. But for SEER cohort had not include the details of surgical procedures, so its' effects on survival and prognosis of BC were not studied and analyzed, which may be a potential factor. In recent years, a number of large randomized trials have shown that radiotherapy could significantly reduce the local recurrence of BC, so as to improve the breast preservation rate and obtain good survival rate³⁸. However, in this study, the radiotherapy rate of BC patients in southern China and SEER cohort were both not high, were 27.18% and 34.22% respectively, and there were significant statistical different between the two ($P=0.001$). Nevertheless, the results of Cox regression analysis showed that radiotherapy had no significant effect on both DFS and OS, which may be related to a low radiotherapy rate, but it should be further studied and analyzed. Besides, chemotherapy was also an important treatment for BC, we found that most of BC patients in southern China are treated with chemotherapy, which was approximately 97.01%. But there were only about 24.08% of BC patients had received

chemotherapy in SEER cohort ($P=0.000$), which indicated that there are great differences in BC treatment between China and the United States.

Additionally, with increasing age (from 70's to 90's), the proportion of young BC patients in SEER cohort increased gradually, while the proportion of elderly patients decreased gradually. Furthermore, the median age at diagnosis was relatively unchanged, they were 36 year, 56 year and 77 year respectively, which may be related to the early diagnosis of BC.

There are some limitations to this study. The data from southern China cohort are from Yunnan Cancer Hospital, Afiliated Tumor Hospital of Guangxi Medical University and The First People's Hospital of Foshan, these data may therefore be slightly different from that of the National Cancer Registry System. In addition, the presence of missing data and limited follow-up time can be considered weaknesses of this study. Other limitations include the lack of data on surgical procedures, KI-67 statuses and endocrine therapy in SEER cohort, which limited the analysis of their influence on patients' survival improvements.

Conclusion

In conclusion, our study suggested that positive lymph node status may cause the difference of morbidity and mortality of BC patients in China. Furthermore, the differences in treatment methods may be the main reason for the differences between China and seer databases.

Data availability

The raw data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Received: 3 November 2021; Accepted: 8 June 2022

Published online: 21 June 2022

References

- Sun, Y. S. *et al.* Risk factors and preventions of breast cancer. *Int. J. Biol. Sci.* **13**, 1387–1397 (2017).
- Cha, Y. J., Kim, E. S. & Koo, J. S. Amino acid transporters and glutamine metabolism in breast cancer. *Int. J. Mol. Sci.* **19**, 25 (2018).
- Zujewski, J. A. *et al.* Knowledge summaries for comprehensive breast cancer control. *J. Glob. Oncol.* **4**, 1–7 (2018).
- Melisko, M. E., Gradishar, W. J. & Moy, B. Issues in breast cancer survivorship: Optimal care, bone health, and lifestyle modifications. *Am. Soc. Clin. Oncol. Educ. Book.* **35**, e22–29 (2016).
- Sun, L., Legood, R., Sadique, Z., Dos-Santos-Silva, I. & Yang, L. Cost-effectiveness of risk-based breast cancer screening programme, China. *Bull. World Health Organ.* **96**, 568–577 (2018).
- Yuan, X. *et al.* Induced abortion, birth control methods, and breast cancer risk: A case-control study in China. *J. Epidemiol.* **29**, 173–179 (2019).
- Lou, Z., Fei, X., Christakos, G., Yan, J. & Wu, J. Improving spatiotemporal breast cancer assessment and prediction in Hangzhou City, China. *Sci. Rep.* **7**, 3188 (2017).
- Peng, Z. *et al.* Treatment and survival patterns of Chinese patients diagnosed with breast cancer between 2005 and 2009 in Southwest China: An observational, population-based cohort study. *Medicine (Baltimore)* **95**, e3865 (2016).
- Hortobagyi, G. N., Edge, S. B. & Giuliano, A. New and important changes in the TNM staging system for breast cancer. *Am. Soc. Clin. Oncol. Educ. Book.* **38**, 457–467 (2018).
- Nahleh, Z. *et al.* Disparities in breast cancer: A multi-institutional comparative analysis focusing on American Hispanics. *Cancer Med.* **7**, 2710–2717 (2018).
- Unlu, O. *et al.* Risk factors and histopathological features of breast cancer among women with different menopausal status and age at diagnosis. *J. BUON* **22**, 184–191 (2017).
- Jarab, M. *et al.* Association of breast cancer grade with response to neoadjuvant chemotherapy assessed postoperatively. *Pol. J. Pathol.* **70**, 91–99 (2019).
- Zhang, X. *et al.* Discussion of relationships among changes of pathological indicators, postoperative lymphedema of the upper limb, and prognosis of patients with breast cancer. *Biosci. Rep.* **39**, 25 (2019).
- Wang, J. *et al.* Value of breast cancer molecular subtypes and Ki67 expression for the prediction of efficacy and prognosis of neoadjuvant chemotherapy in a Chinese population. *Medicine (Baltimore)* **95**, e3518 (2016).
- Wang, X. *et al.* Prevalence of BRCA1 and BRCA2 gene mutations in Chinese patients with high-risk breast cancer. *Mol. Genet. Genom. Med.* **7**, e677 (2019).
- Song, J. L., Chen, C., Yuan, J. P. & Sun, S. R. Progress in the clinical detection of heterogeneity in breast cancer. *Cancer Med.* **5**, 3475–3488 (2016).
- Tesarova, P. Specific aspects of breast cancer therapy of elderly women. *Biomed. Res. Int.* **2016**, 1381695 (2016).
- Barrios, C. H., Reinert, T. & Werutsky, G. Global breast cancer research: Moving forward. *Am. Soc. Clin. Oncol. Educ. Book.* **38**, 441–450 (2018).
- DeSantis, C. E., Ma, J., Goding Sauer, A., Newman, L. A. & Jemal, A. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA Cancer J. Clin.* **67**, 439–448 (2017).
- Sung, H. *et al.* The impact of breast cancer-specific birth cohort effects among younger and older Chinese populations. *Int. J. Cancer* **139**, 527–534 (2016).
- Wu, A. H., Vigen, C., Lee, E., Tseng, C. C. & Butler, L. M. Traditional breast cancer risk factors in Filipina Americans compared with Chinese and Japanese Americans in Los Angeles County. *Cancer Epidemiol. Biomark. Prev.* **25**, 1572–1586 (2016).
- Holm, J. *et al.* Risk factors and tumor characteristics of interval cancers by mammographic density. *J. Clin. Oncol.* **33**, 1030–1037 (2015).
- Goldhirsch, A. *et al.* Strategies for subtypes—dealing with the diversity of breast cancer: Highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann. Oncol.* **22**, 1736–1747 (2011).
- Wang, K. *et al.* Comparison of clinicopathological features and treatments between young (≤ 40 years) and older (> 40 years) female breast cancer patients in West China: A retrospective, epidemiological, multicentre, case only study. *PLoS One* **11**, e0152312 (2016).
- Feng, F. *et al.* Comparison of epidemiological features, clinicopathological features, and treatments between premenopausal and postmenopausal female breast cancer patients in western China: A retrospective multicenter study of 15,389 female patients. *Cancer Med.* **7**, 2753–2763 (2018).
- Huang, Z. *et al.* Breast cancer incidence and mortality: Trends over 40 years among women in Shanghai, China. *Ann. Oncol.* **27**, 1129–1134 (2016).

27. Ye, D. M. *et al.* Clinical and epidemiologic factors associated with breast cancer and its subtypes among Northeast Chinese women. *Cancer Med.* **8**, 7431–7445 (2019).
28. Radecka, B. & Litwiniuk, M. Breast cancer in young women. *Ginekol. Pol.* **87**, 659–663 (2016).
29. Bhattacharyya, G. S. *et al.* Overview of breast cancer and implications of overtreatment of early-stage breast cancer: An Indian perspective. *JCO Glob. Oncol.* **6**, 789–798 (2020).
30. Martei, Y. M., Pace, L. E., Brock, J. E. & Shulman, L. N. Breast cancer in low- and middle-income countries: Why we need pathology capability to solve this challenge. *Clin. Lab. Med.* **38**, 161–173 (2018).
31. Tian, Z. *et al.* Atypical ubiquitin-binding protein SHARPIN promotes breast cancer progression. *Biomed. Pharmacother.* **119**, 109414 (2019).
32. Zhang, L. *et al.* Comparison of breast cancer risk factors among molecular subtypes: A case-only study. *Cancer Med.* **8**, 1882–1892 (2019).
33. Ding, L., Zhang, Z., Xu, Y. & Zhang, Y. Comparative study of Her-2, p53, Ki-67 expression and clinicopathological characteristics of breast cancer in a cohort of northern China female patients. *Bioengineered* **8**, 383–392 (2017).
34. Fisusi, F. A. & Akala, E. O. Drug combinations in breast cancer therapy. *Pharm. Nanotechnol.* **7**, 3–23 (2019).
35. Zhang, M. *et al.* Expression of PD-L1 and prognosis in breast cancer: a meta-analysis. *Oncotarget* **8**, 31347–31354 (2017).
36. Bao, X. *et al.* Present and changing trends in surgical modalities and neoadjuvant chemotherapy administration for female breast cancer in Beijing, China: A 10-year (2006–2015) retrospective hospitalization summary report-based study. *Thorac. Cancer* **9**, 707–717 (2018).
37. Bartelink, H. The changing landscape in radiotherapy for breast cancer: Lessons from long term follow-up in some European breast cancer trials. *Radiother. Oncol.* **121**, 348–356 (2016).
38. Gupta, A., Ohri, N. & Haffty, B. G. Hypofractionated radiation treatment in the management of breast cancer. *Expert Rev. Anticancer Ther.* **18**, 793–803 (2018).

Author contributions

S.C.T. designed the study, F.Z. collected and interpreted the data, Y.J.N. and S.C.T. carried out the statistical analysis, R.G., H.Y.L. and D.Q.L. provided administrative support, Y.J.N., D.C., and W.L.C. provision of study materials or patients. All authors read and approved the final manuscript.

Funding

The present study was supported by the National Natural Science Foundation of China (Grant no. 81960542, 81960517, 81660437, 81760478); Scientific Research Foundation of Yunnan Education Department (Grant no. 2019J1288 and 2020J0198) and Joint project of Kunming Medical University (2018FE001(-068)).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-14573-4>.

Correspondence and requests for materials should be addressed to R.G., H.L. or S.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022