Analysis and prediction of second primary malignancy in patients with breast cancer

QUANYI LONG[1,2], FEILONG ZHAO[3] and HONGJIANG LI[1]

[1]Department of Breast Surgery, West China Hospital, Sichuan University, Chengdu, Sichuan 610041, P.R. China; [2]International College, Krirk University, Bangkok 10220, Thailand; [3]3D Medicines, Inc., Shanghai 201100, P.R. China

Received June 11, 2022; Accepted October 6, 2022

DOI: 10.3892/mco.2022.2593

Abstract. Second primary malignancy (SPM) is common in breast cancer (BC). The present study aimed to profile the characteristics of BC with SPM and to identify patients at high risk of SPM. Clinical and outcome data of BC cases were retrieved from the SEER database. Principal component analysis and a random forest model were utilized to create a model for predicting the occurrence of SPMs. Of the 286,047 BC cases analyzed, 9.32% developed SPMs. Approximately 70% of BCs that developed SPMs were ductal carcinoma and 71% of BCs that developed SPMs were human epidermal growth factor receptor 2 $(HER2)^-$/hormone receptor $(HR)^+$. The overall survival (OS) of the SPM cohort was significantly worse (hazard ratio: 1.49; 95% CI: 1.44-1.53; log-rank P<0.001). After adjusting for metastasis status, SPM was still a poor prognostic factor (hazard ratio: 1.71; 95% CI: 1.70-1.82; log-rank P<0.001). Of note, 50.5% of the SPMs occurred in the breast and the OS of the breast SPM group was significantly better than that of the other single-organ SPM group (hazard ratio: 0.46; 95% CI: 0.45-0.49; log-rank P<0.001) and the multiple-organ SPM group (hazard ratio: 0.44; 95% CI: 0.39-0.50; log-rank P<0.001). A random forest model created from clinical features predicted SPM with a positive predictive value of 32.3% and negative predictive value of 90.7% in the testing set. Thus, SPM occurs in nearly 1/10 of BC survivors and its existence and occurrence site significantly influence OS. SPM may be partly predicted from clinical features. In addition, it was indicated that postmenopausal elderly patients with a $HER2^-/HR^+$ molecular subtype should be more watchful and undergo screenings for SPMs.

Correspondence to: Professor Hongjiang Li, Department of Breast Surgery, West China Hospital, Sichuan University, No. 37 Guoxue Alley, Wuhou, Chengdu, Sichuan 610041, P.R. China
E-mail: lihongjiang@sohu.com

Key words: second primary malignancy, prevalence, site, survival, prediction

Introduction

Breast cancer (BC) is the most prevalent cancer type among females worldwide (1). Advances in early systematic screening, effective treatments and supportive care have significantly prolonged the survival time of patients with BC (2). Due to the high incidence and good prognosis of BC, the risk of developing a second primary malignancy (SPM) thereafter may turn into a serious health issue both for the patients and health care system. A large study in the US found a higher risk of SPM associated with BC than with other cancers in women (3).

An SPM is a second, unrelated cancer in a person who has previously experienced another cancer at any time. The exact incidence of SPMs is uncertain, though studies have provided certain insight. One study evaluated over 2 million people who developed the 10 most common types of cancer from 1992 to 2008, and >10% of them developed an SPM (4). Previous population-based research has examined the risk of developing SPMs among initial primary BC survivors compared to the general population, but the results from these studies were inconsistent in their risk estimation, finding a wide risk range of 15-45% for any type of SPM (2). Therefore, accurately estimating the SPM risk and profiling the characteristics of patients at risk would be valuable.

An SPM may occur in the same tissue or organ as the first cancer or in another region of the body (5). These second cancers may be related to a genetic predisposition, common risk factors or treatments for the original cancer, or they may simply occur sporadically, as cancer commonly does (6,7). The link between the characteristics of the primary cancer and the risk of SPM is controversial (8,9). Little is known regarding the simultaneous effect of intrinsic factors, such as age at diagnosis, sex, marital status, ethnicity, hormone receptor (HR) status and tumor characteristics, which lead to the development of a new malignancy in a BC survivor.

The present study aimed to comprehensively profile the characteristics of patients with BC harboring an SPM and to further identify patients at high risk of developing SPMs using a large, population-based cohort. First, the demographic, clinical and histological differences between patients with BC with only one primary malignancy (OOPM) and with SPMs were retrieved. The influence of SPM on prognosis was then investigated. Finally, the intrinsic factors associated with the

development of SPM were evaluated and a machine learning model was established to identify BC survivors who were at high risk of developing SPMs.

Patients and methods

Data sources. Data were extracted from the Surveillance, Epidemiology and End Results (SEER) research database. The SEER program is the most authoritative and premier source of cancer statistics in the US, collecting demographics, tumor characteristics and survival data. The SEER data of the version November 2019 (https://seer.cancer.gov/data-software/documentation/seerstat/nov2019/) were downloaded. The downloaded data contained four compressed files: file-1Bvek1.rar, file9c7Clt.rar, fileb0bhOt.rar and fileYhNIxM.rar. Each of the above files contained two files, titled xxx.dic and xxx.txt. The dic file is the column id of the txt file is the content. An in-house python script was used to read and process these files. In brief, a function named extract_colname was used to extract the column names from the dic file, which were assigned to the corresponding txt file. The patients with BC were then selected and the histology was rated according to criteria in the 3rd edition of the International Classification of Diseases for Oncology (10).

Study population. Patients with BC diagnosed after 2010 were included because molecular subtypes were available from then on. Only cases with complete data, without missing values on important covariates (age, ethnicity, tumor site, grade, size) were eligible. Cases that were reported from a death certificate or autopsy were excluded and a 2-month latency exclusion was set to further distinguish SPMs from simultaneous cancers. The identified patients with BC were then categorized into two cohorts: The OOPM cohort and the SPM cohort. The study design and workflow are presented in Fig. 1.

Definition of SPM. According to the SEER rules for classifying multiple primary cancers, the definition was dependent on the cancer site of origin, date of diagnosis, histology, tumor behavior (i.e., *in situ* vs. invasive) and laterality of any paired organs. In general, all SPMs occurring 2 or more months after the initial diagnosis were considered separate primary cancers unless the medical record stated that the tumor was recurrent or metastatic (11,12). There were also two key variables indicating multiple primary malignancies in SEER, the 'total number of *in situ*/malignant tumors for patient' and the 'sequence number' of the multiple primary malignancies. The former was used to identify patients with an SPM and the latter to index the sequence of multiple malignancies. A random variable that was named 'indicator of SPM' was defined to indicate whether the patient had developed one.

Model creation using unsupervised and supervised methods. An unsupervised machine learning method called factor analysis of mixed data (FAMD), which is generally used to analyze datasets containing both quantitative and qualitative variables, was used to transform data. In brief, FAMD may be regarded as a mix between principal component analysis (PCA) and multiple correspondence analysis. It acts as PCA for quantitative variables and as multiple correspondence analysis for qualitative variables. This was achieved by an R package FactoMineR (13).

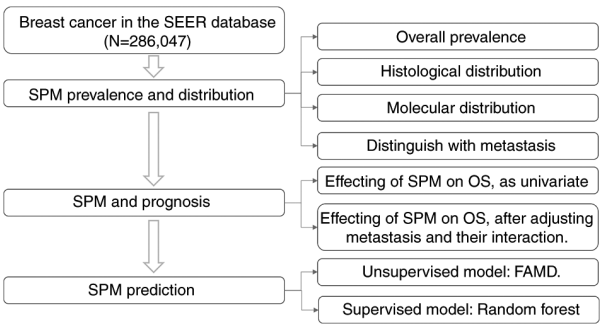

Figure 1. Study design and workflow of the present study. OS, overall survival; SPM, second primary malignancy; SEER, Surveillance, Epidemiology and End Results; FAMD, factor analysis of mixed data.

The total dataset was randomly split 75/25% into the training set and testing set, stratified by the existence of SPMs. A popular supervised machine learning method called random forest was applied to the training set to predict the likelihood of developing SPMs. The performance of the random forest classifier was evaluated in the testing set with 50 repetitions to reduce the influence of randomization. The outcome was visualized by the receiver operating characteristic (ROC) curve. Feature importance was calculated by their contribution to the prediction ability of the model. The above steps were implemented in python using the sci-kit-learn package.

Statistical analysis. To compare distributions between variables, the χ^2 test was generally applied for discrete variables, Student's t-test for continuous variables satisfying a normal distribution and the Mann-Whitney U-test for continuous variables otherwise. For survival analysis, both the non-parametric Kaplan-Meier model and the semi-parametric Cox proportional hazard model were used to evaluate the influence of variables on overall survival (OS); when both methods produced a significant P-value, the result was regarded to be significant. $P<0.05$ was considered to indicate statistical significance.

Results

Clinical, histological and molecular characteristics of patients with BC with SPMs. A total of 286,047 patients with BC were identified from the SEER database. Of them, 26,657 (9.32%) developed SPMs within a maximum follow-up of ~7 years. The characteristics of the patients with OOPM and SPM are compared in Table I. In general, ~99% of patients with BC were females and ~50% of BCs were in the left breast. Specifically, the SPM cohort had significantly more patients with well (23.91%) or moderately (44.22%) differentiated pathology and with more widowed patients (14.87%), while the OOPM cohort had a significantly higher percentage of patients poorly differentiated tumors (31.86%) and of married patients (55.04%). The SPM cohort was significantly older than the OOPM cohort (median age, 63 vs. 60 years; $P<0.001$). Of note, the SPM frequency was significantly ($P=0.003$) higher in stage M0 (25,316/245,406, 10.32%) than in stage M1 (1,089/11,590, 9.39%; Table I).

Table I. Characteristics of patients with BC with OOPM and SPM.

Parameter	OOPM (n=259,390)	SPM (n=26,657)
Sex		
Female	257,517 (99.28)	26,432 (99.16)
Male	1,873 (0.72)	225 (0.84)
Age at diagnosis, years [median (range)]	60.0 (2.0-117.0)	63.0 (21.0-103.0)
Marital status		
Married	142,757 (55.04)	14,109 (52.93)
Single	39,001 (15.04)	3,937 (14.77)
Widowed	32,868 (12.67)	3,964 (14.87)
Divorced	27,488 (10.60)	2,910 (10.92)
Separated	2,822 (1.09)	257 (0.96)
Unmarried or domestic partner	753 (0.29)	93 (0.35)
Laterality		
Left	131,723 (50.78)	13,197 (49.51)
Right	127,288 (49.07)	13,444 (50.43)
Left or right[a]	49 (0.02)	2 (0.01)
Bilateral	42 (0.02)	0 (0.00)
Ethnicity		
White	202,991 (78.26)	21,556 (80.86)
Black	29,453 (11.35)	2,765 (10.37)
Asian or Pacific Islander	23,403 (9.02)	2,063 (7.74)
American Indian/Alaska Native	1,570 (0.61)	159 (0.60)
Grade		
I: Well differentiated	55,797 (21.51)	6,373 (23.91)
II: Moderately differentiated	108,397 (41.79)	11,788 (44.22)
III: Poorly differentiated	82,637 (31.86)	7,227 (27.11)
IV: Undifferentiated	900 (0.35)	74 (0.28)
BC subtype		
HER2+/HR+	26,825 (10.34)	2,199 (8.25)
HER2+/HR-	11,292 (4.35)	876 (3.29)
HER2-/HR+	177,409 (68.39)	19,487 (73.10)
Triple negative	28,180 (10.86)	2,402 (9.01)
T stage		
T1	146,231 (56.37)	14,865 (55.76)
T2	76,558 (29.51)	7,865 (29.50)
T3	14,686 (5.66)	1,721 (6.46)
T4	6,170 (2.38)	697 (2.61)
N stage		
N0	170,761 (65.83)	17,547 (65.83)
N1	60,316 (23.25)	6,092 (22.85)
N2	13,472 (5.19)	1,408 (5.28)
N3	10,488 (4.04)	1,177 (4.42)
M stage		
M0	245,406 (94.61)	25,316 (94.97)
M1	11,590 (4.47)	1,089 (4.09)
Stage		
I	120,561 (46.48)	12,247 (45.94)
II	94,164 (36.30)	9,664 (36.25)
III	29,340 (11.31)	3,268 (12.26)
IV	11,590 (4.47)	1,089 (4.09)

[a]Only one side involved, right or left but unspecified. Values are expressed as n (%) unless otherwise specified. SPM, second primary malignancy; OOPM, only one primary malignancy; BC, breast cancer; HR, hormone receptor; HER2, human epidermal growth factor receptor 2.

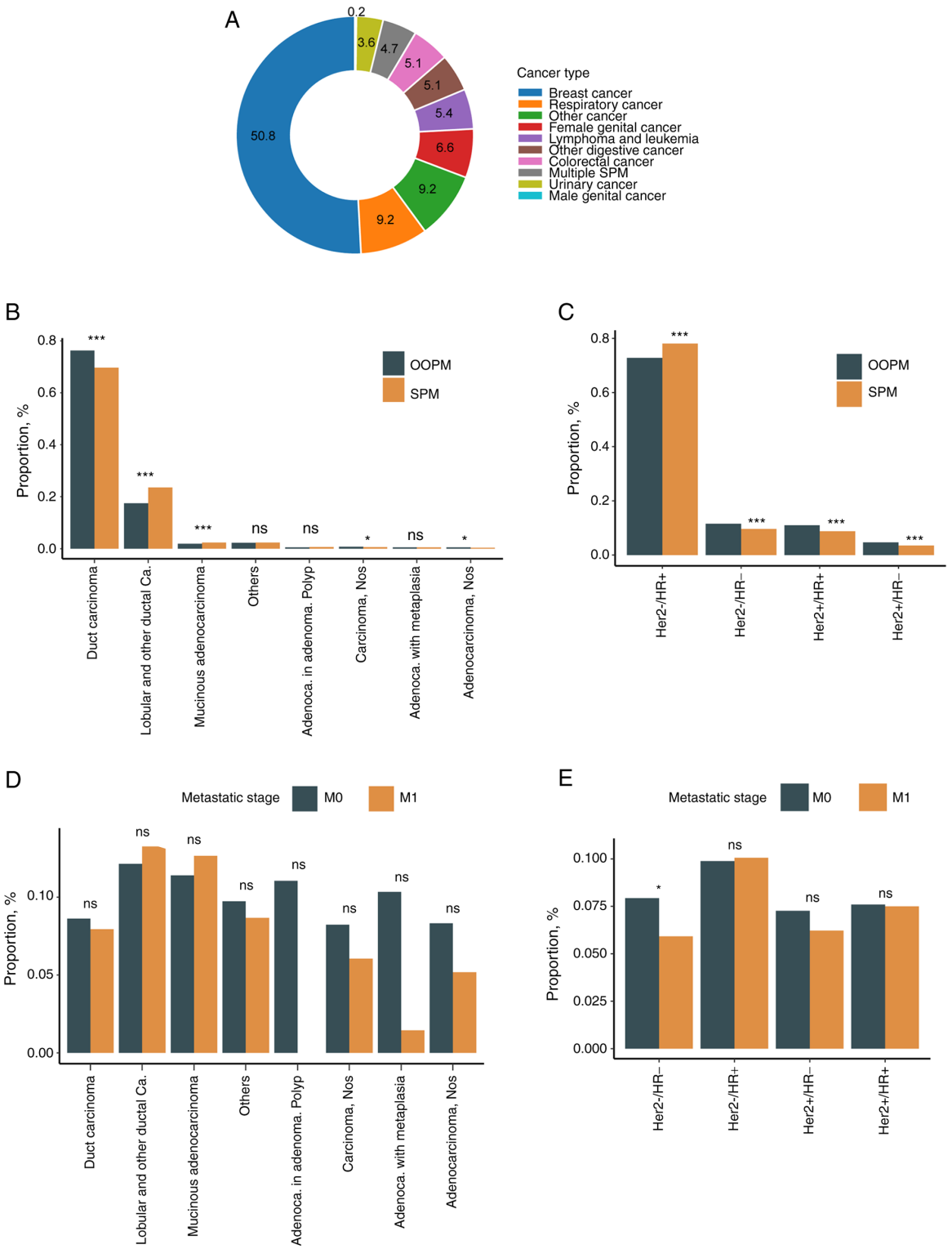

Figure 2. Characteristics of patients with SPM. (A) Distribution of cancer types among SPMs. (B) Distribution of cancer types in the OOPM and SPM cohorts. The enrichment of SPMs in lobular/ductal carcinoma is evident. (C) Distribution of molecular subtypes in the OOPM and SPM cohorts. The enrichment of SPMs among the $HER2^{-}/HR^{+}$ molecular subtype is evident. (D) Incidence rates of SPMs in each histological type in the nonmetastatic cohort and the metastatic cohort. There was no significant difference in the incidence rate of SPMs between the metastatic and nonmetastatic cohorts. (E) Incidence rates of SPMs in each molecular subtype in the nonmetastatic cohort and the metastatic cohort. The incidence rate of SPMs in the $HER2^{-}/HR^{+}$ molecular subtype in the nonmetastatic cohort was significantly higher than that in the metastatic cohort. $^{*}P<0.05$; $^{***}P<0.001$; ns, no significance. SPM, second primary malignancy; OOPM, only one primary malignancy; HR, hormone receptor; HER2, human epidermal growth factor receptor 2; Ca, carcinoma; Nos, not otherwise specified.

The cancer types of the SPMs are profiled in Fig. 2A. Half of the SPMs occurred in the breast and half were found evenly in the other systems, including the respiratory, female germline, lymphatic/leukocytic, colorectal, other digestive and urinary systems. of note, 0.2% of SPMs were detected in the male germline system.

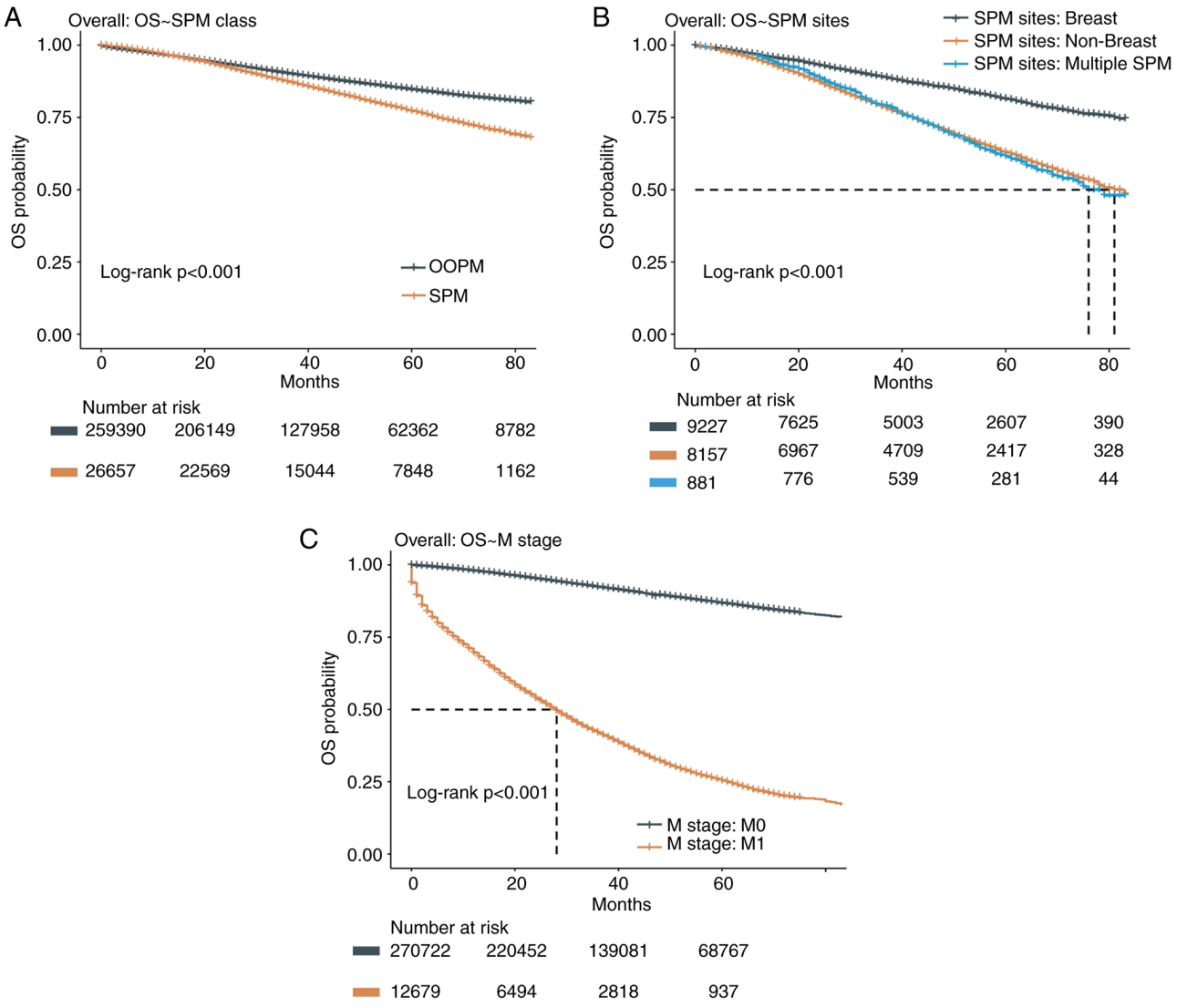

Figure 3. Survival analysis of SPMs against OS. (A) Impact of SPM on OS. The OS of the SPM cohort was significantly worse than that of the OOPM cohort. (B) Effect of SPM location on OS. The OS of patients with SPMs in the breast was significantly better than that of patients SPMs in other organs. (C) Effect of metastasis status on OS. The OS of patients in stage M1 was significantly inferior to that of patients in stage M0. SPM, second primary malignancy; OOPM, only one primary malignancy; OS, overall survival.

The histological type distribution in the SPM cohort was compared with that in the OOPM cohort (Fig. 2B). Most SPMs were ductal carcinoma in both cohorts, but the frequency of ductal carcinoma was significantly lower in the SPM cohort (69.1%) than in the OOPM cohort (75.2%, Fisher's exact P<0.001). On the other hand, the frequency of lobular carcinoma, which was the second most common carcinoma in both cohorts, was significantly more common in the SPM cohort (23.3%) than in the OOPM cohort (17.1%, Fisher's exact P<0.001).

Molecular status had been determined by a combination of immunohistochemistry, fluorescence *in situ* hybridization, chromogenic *in situ* hybridization and other methods by the SEER group (https://seer.cancer.gov/seerstat/databases/ssf/her2-derived.html). The $HER2^-/HR^+$ subtype was significantly enriched in the SPM cohort compared to the OOPM cohort (78.1% vs. 72.8%, P<0.001; Fig. 2C). Of note, $HER2^-/HR^+$ was the most common subtype in both the OOPM cohort and SPM cohort, and the $HER2^-/HR^+$ subtype was more likely to have an SPM (10%) than other subtypes. The proportion of $HER2^+/HR^+$, $HER2^+/HR^-$ and triple-negative subtypes was generally lower in the SPM cohort than in the OOPM cohort.

Since it is at times difficult to distinguish metastasis and SPM, the SPM frequency was compared between stage M0 and stage M1, stratified by histological and molecular subtype. Although the SPM frequency was slightly higher in stage M0 than in stage M1 across numerous histological types, no significant difference in SPM frequency was detected in any histological type (Fig. 2D). As for molecular subtypes, the SPM frequency was significantly higher in stage M0 in the $HER2^-/HR^-$ subtype (Fig. 2E).

OS of patients with BC with SPMs. The OS of the SPM cohort was significantly worse than that in the OOPM cohort (hazard ratio: 1.49; 95% CI: 1.44-1.53; log-rank P<0.001), indicating the role of SPMs in accelerating patient death (Fig. 3A). As more than half of SPMs occurred in the breast again, while the rest seemed to occur in other organs at random, their OS rates

Table II. Univariate and multivariate Cox regression analysis of SPM group and metastasis status for overall survival.

	Univariate		Multivariate	
Factor	Hazard ratio (95% CI)	P-value	Hazard ratio (95% CI)	P-value
SPM (vs. OOPM)	1.49 (1.44-1.53)	<0.001	1.71 (1.70-1.82)	<0.001
Metastasis status (M1 vs. M0)	11.37 (11.09-11.67)	<0.001	12.64 (12.38-13.05)	<0.001
Interaction term[a]			0.40 (0.36-0.43)	<0.001

[a]Interaction between SPM group and metastasis status. SPM, second primary malignancy; OOPM, only one primary malignancy.

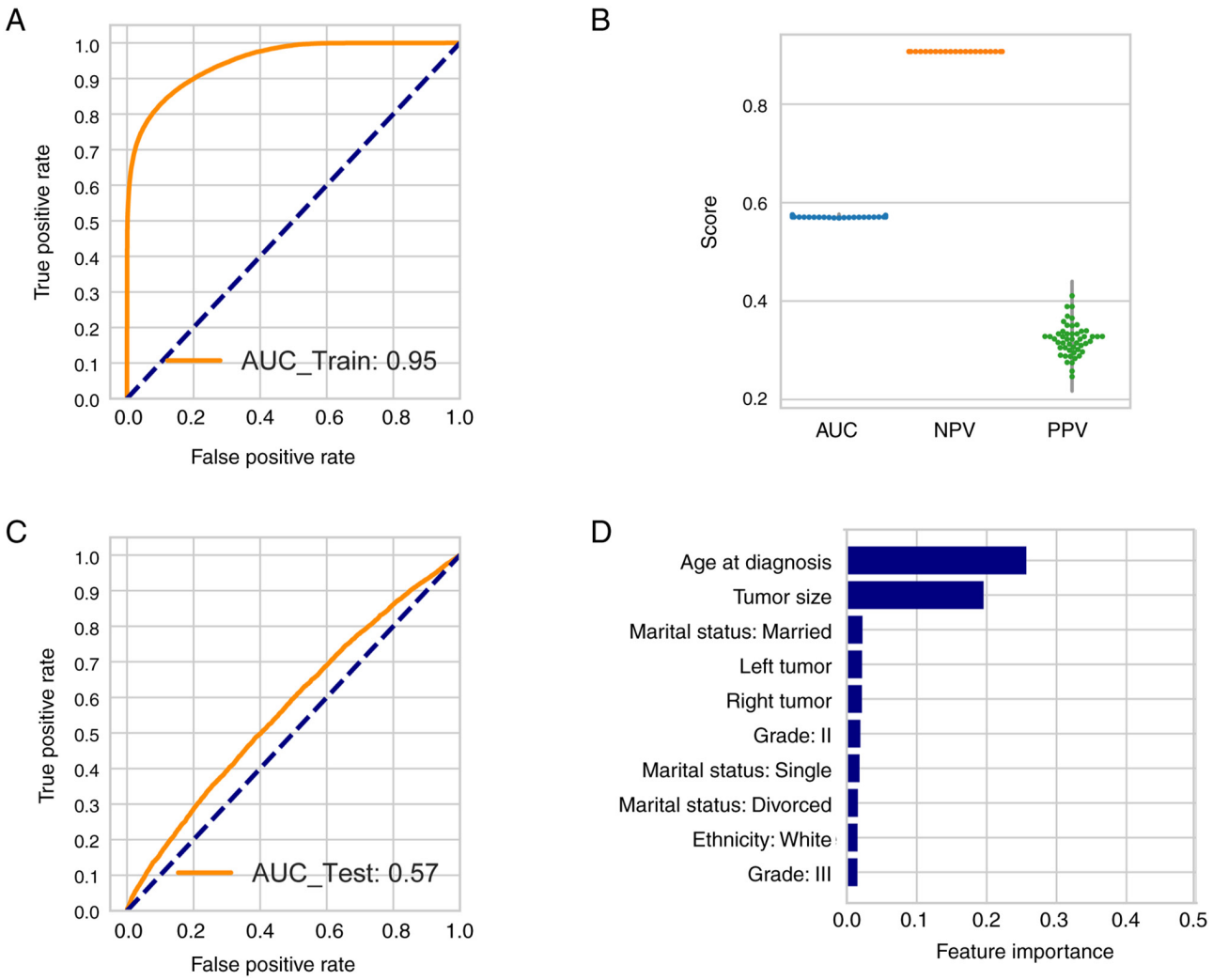

Figure 4. Performance of the model in predicting the occurrence of SPMs. (A) ROC curve illustrating the diagnostic accuracy of the model in the training set. (B) AUC, PPV and NPV of the model in testing set after 50 repeats. The mean AUC, NPV and PPV in the testing set were 0.57, 0.91 and 0.32, respectively. (C) A representative ROC curve in the testing set. (D) The top 10 features that contributed to the performance of the model in the testing set. SPM, second primary malignancy; ROC, receiver operating characteristic; AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value.

were compared to explore the influence of SPM location on survival. The OS of breast SPMs was significantly better than that of other single-organ SPMs (hazard ratio: 0.46; 95% CI: 0.45-0.49; log-rank P<0.001) and that of multiple-organ SPMs (hazard ratio: 0.44; 95% CI: 0.39-0.50; log-rank P<0.001; Fig. 3B), while there was no significant OS difference between the latter two groups (hazard ratio: 0.98; 95% CI: 0.87-1.10; log-rank P=0.69). Median OS was not achieved (i.e. the survival rate remained >50%) in patients with SPMs of the breast, while it was 81 and 76 months in patients with SPMs of other organs and in patients with multiple SPMs, respectively.

Since distant metastasis is a key factor influencing OS, this was validated in the present dataset (Fig. 3C). To better understand the role of distant metastasis and SPM on OS,

multivariate Cox survival analysis was performed for distant metastasis, SPM and their interacting effect as covariates. The results indicated that both distant metastasis and SPM were significantly associated with OS after adjusting for each other (for SPM, hazard ratio: 1.71; 95% CI: 1.70-1.82; for M1, hazard ratio: 12.64, 95% CI: 12.38-13.05; Table II). The significant interaction P-value implies that the influence of SPM on OS was more significant in patients with stage M0.

Predicting the occurrence of SPMs in patients with BC. To detect whether the SPM and OOPM cohorts may be distinguished by certain features, unsupervised transformation was performed using FAMD, which was an extension of PCA. The general purpose of PCA is to find transformed features that may cluster the patients into two or more clusters and the transformed features are a combination of the original variables. In the present study, there were 23 original variables. After transformation, the top 5 features were extracted. Only slightly >10% of the variance of the data was able to be explained by the top five transformed features (Fig. S1A), while the top two transformed features were only able to explain 6.5% of the variance. The tumor stage contributed the most to the variance detected by the top two transformed features (Fig. S1B). Patients with SPMs were not able to be clustered together using the top two transformed features (Fig. S1C). Therefore, using supervised learning, a random forest model was created to predict the probability of SPM in patients with BC.

The patient population was randomly split into a training set (75%) and a testing set (25%), each stratified by the presence of SPMs. Parameters including maximum depth and class weight were learned from the training set and the parameters that generated the highest positive predictive value (PPV) in the out-of-bag mode were adopted to create the random forest model. The model generated an overall area under the curve (AUC) of 0.95 in the training set (Fig. 4A). To reduce the influence of randomization, the model was tested 50 times using different random seeds. The mean consistency of the testing set was 0.91, but in the imbalance dataset, the PPV and negative predictive value (NPV) were more important features. The mean AUC, NPV and PPV in the testing set were 0.57, 0.91 and 0.32, respectively (Fig. 4B). The median value of the above parameters was the same as their mean value.

A representative ROC curve with its AUC in the testing set is illustrated in Fig. 4C. The top 10 features contributing to the estimation in the testing set were displayed in Fig. 4D. Age at diagnosis and tumor size had the highest weight, at ~25 and 20%, respectively.

Discussion

BC has the highest incidence among all cancers in the world in females, since its incidence has surpassed that of lung cancer (14). Most studies have indicated that the clinical factors influencing the survival of patients with BC include tumor stage, estrogen receptor status, progesterone receptor status and HER2 status (15). However, with the aging of the population and the continuous extension of the survival time of patients with BC, the incidence rate of multiple primary malignancies has gradually increased in recent years (16). The OS of patients with BC is related to not only the primary malignancy but also the nature of SPMs and the organs bearing the SPMs. Determining the risk of SPMs in patients with BC, predicting patients at high risk for SPMs, and closely monitoring these patients have a vital role in improving the OS and guiding clinical practice.

Carcinogenesis is a multistep, long-term process. As life expectancy increases, the likelihood of being diagnosed with cancer also increases. Therefore, the risk of developing SPMs gradually increases with age (17,18). In addition, compared with the general population, cancer patients have a much higher risk of developing SPMs (16). The results of the present study indicated that 9.32% of patients with BC developed SPMs within 7 years after the diagnosis of the primary malignancy and that this proportion would keep increasing if the follow-up were to be continued. This finding is similar to the results of Xiao *et al* (16).

The present study suggested that approximately half of SPMs in patients with BC occurred in the breast, while the rest appeared to occur randomly in other organs. This finding suggests that the primary malignancy of BC may change the mammary gland microenvironment and contribute to the occurrence of SPMs (19), or it may be interpreted through the notion of hereditary cancer syndromes reported in previous studies. Hereditary BC and ovarian cancer syndrome are hereditary malignancies that may be confirmed by detecting germline mutations in the *BRCA1* or *BRCA2* genes (20). Compared with the general population, females with *BRCA1* or *BRCA2* gene mutations have a significantly higher risk of BC and ovarian cancer (21). In the present study, SPM occurred most frequently in the breast, which reflects the susceptibility of the breast to the invasion of primary BC, in line with the studies mentioned above, which suggested that the implementation of preventive mastectomy may obtain a survival benefit for patients with BC. Numerous studies on the association between BC and colon cancer have reported the coexistence of common extrinsic and genetic predisposition factors (18). A prospective cohort study of a female BC population suggested that the standardized incidence ratio of secondary primary colorectal cancer in BC survivors was 1.59 (22). *BRCA* mutations may increase the risks of colorectal cancer (23), ovarian cancer, pancreatic cancer and prostate cancer (21). The present study indicated that nearly half of the SPMs occurred randomly in organs other than the breast (such as the digestive, respiratory, blood and reproductive systems). The random occurrence of SPMs may reflect the genetic tendency of these patients (24), which may also correspond to the more common sites mentioned in certain studies (18,25-27).

Although a small number of studies have looked into whether menopausal women are more likely to develop SPMs, the present study found that patients with SPMs were mostly $HER2^-/HR^+$ menopausal patients with a median age of 63 years, consistent with the finding of Xiao *et al* (16) that >70% of the patients had reached menopause prior to the diagnosis of the SPMs. Therefore, it may be reasonable to suggest that SPM monitoring should begin after the end of the BC regimen. Postmenopausal elderly patients with a $HER2^-/HR^+$ molecular subtype should be more watchful for SPMs. In particular, for patients who deny a family history of BC at the first diagnosis

but a hereditary tumor-related syndrome is detected during the follow-up, as well as in patients with a known family history of BC, ovarian cancer, pancreatic cancer, prostate or gastrointestinal cancer, not only routine reexaminations should be performed after BC operation to exclude recurrence and metastasis, but also the family tumor history of patients should be reviewed at each reexamination, so as to avoid missing a diagnosis of SPM due to ignoring hereditary tumor syndrome. Misdiagnosis or missed diagnosis should be avoided and early detection, early diagnosis and early treatment should be aimed for.

The OS of the SPM cohort was significantly lower than that of the OOPM cohort, indicating that the occurrence of SPM had a certain role in accelerating disease progression and deterioration. Compared with the patients with SPMs in non-breast organs, the patients with SPMs in the breast had significantly better OS. Compared with patients with SPMs in non-breast organs and patients with multiple SPMs, the patients with SPMs in the breast had 54 and 56% lower risks of death, respectively. In addition, OS was not significantly different between patients with SPMs in non-breast organs and patients with multiple SPMs, which indicates that the organs bearing SPMs had a significantly greater impact on prognosis than other factors, such as the number of SPMs.

To predict the occurrence of SPM at the time when the primary BC was diagnosed, a supervised machine learning model was created based on clinical characteristics and features of the primary tumor, such as age at diagnosis, marital status and tumor location. The model of the present study had a PPV of 32% and NPV of 91%. This performance is not very good, but this was the best result that was achieved when using the above features after comparing various models. Compared to the unsupervised machine learning model, which was not able to clearly distinguish SPMs from OOPMs, the model of the present study achieved an acceptable PPV and a high NPV. Of all the features used to create the model, age at diagnosis and tumor size were the two most important features predicting SPM, which is reasonable and consistent with previous reports (4,28). It may be possible to further improve the performance of the model by adding more features, such as genetic variation.

The present study has several limitations. First, it is retrospective and the data originated from different centers; therefore, it has limitations inherent to such studies such as heterogeneity regarding data recording and patient management etc. Furthermore, the differential diagnosis between SPMs and metastatic lesions is still difficult, so diagnostic confusion between the two is inevitable. Finally, there is a lack of information regarding the treatment given after surgery or diagnosis, which is an important prognostic variable. However, considering the large population base, the present study made valuable contributions.

In conclusion, the present study describes the clinical, histological and molecular characteristics of patients with BC with SPMs based on the SEER dataset. The results suggested that the OS of the SPM cohort was significantly worse than that of the OOPM cohort, and the OS of the patients with SPMs in the breast was significantly better than that of the patients with SPMs in other organs. Furthermore, the negative effect of SPM on OS was independent of the metastasis status. A supervised machine learning model was created that had a 32% PPV and 91% NPV using certain clinical characteristics and characteristics of the primary malignancy. In addition, postmenopausal elderly patients with a $HER2^-/HR^+$ molecular subtype should be more watchful for SPMs. The present results suggest that SPMs in the breast should be considered a prognostic factor; the association between BC and SPMs should not be ignored only because of metastasis. In addition, adequate diagnosis and long-term regular follow-up are of great significance to patients with malignancies. Therefore, attention should be paid to SPM monitoring to avoid misdiagnoses or missed diagnoses and to achieve early detection, early diagnosis and early treatment in these patients.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

All data used in this study are available from the SEER research database.

Authors' contributions

Conception and design: QL and HL. Collection and collation of data: QL and FZ. Data analysis and interpretation: FZ. Manuscript writing: All authors. All authors have read and approved the final manuscript. Data authentication is not applicable.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Global Burden of Disease Cancer Collaboration; Fitzmaurice C, Akinyemiju TF, Al Lami FH, Alam T, Alizadeh-Navaei R, Allen C, Alsharif U, Alvis-Guzman N, Amini E, *et al*: Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016: A systematic analysis for the global burden of disease study. JAMA Oncol 4: 1553-1568, 2018.
2. Molina-Montes E, Requena M, Sanchez-Cantalejo E, Fernandez MF, Arroyo-Morales M, Espin J, Arrebola JP and Sánchez MJ: Risk of second cancers cancer after a first primary breast cancer: A systematic review and meta-analysis. Gynecol Oncol 136: 158-171, 2015.
3. Song F, Qureshi AA, Giovannucci EL, Fuchs CS, Chen WY, Stampfer MJ and Han J: Risk of a second primary cancer after non-melanoma skin cancer in white men and women: A prospective cohort study. PLoS Med 10: e1001433, 2013.

4. Donin N, Filson C, Drakaki A, Tan HJ, Castillo A, Kwan L, Litwin M and Chamie K: Risk of second primary malignancies among cancer survivors in the United States, 1992 through 2008. Cancer 122: 3075-3086, 2016.
5. Heard A, Roder D and Luke C: Multiple primary cancers of separate organ sites: Implications for research and cancer control (Australia). Cancer Causes Control 16: 475-481, 2005.
6. Hemminki K and Boffetta P: Multiple primary cancers as clues to environmental and heritable causes of cancer and mechanisms of carcinogenesis. IARC Sci Publ 2004: 289-297, 2004.
7. Travis LB, Rabkin CS, Brown LM, Allan JM, Alter BP, Ambrosone CB, Begg CB, Caporaso N, Chanock S, DeMichele A, *et al*: Cancer survivorship-genetic susceptibility and second primary cancers: Research strategies and recommendations. J Natl Cancer Inst 98: 15-25, 2006.
8. Bergfeldt K, Einhorn S, Rosendahl I and Hall P: Increased risk of second primary malignancies in patients with gynecological cancer. A Swedish record-linkage study. Acta Oncol 34: 771-777, 1995.
9. Lee KD, Chen CY, Huang HJ, Wang TY, Teng D, Huang SH, Lai CH and Chen MC: Increased risk of second primary malignancies following uterine cancer: A population-based study in Taiwan over a 30-year period. BMC Cancer 15: 393, 2015.
10. Fritz, April, Percy, Constance, Jack, Andrew, Shanmugaratnam, Kanagaratnam, Sobin, Leslie H, *et al*: International classification of diseases for oncology, 3rd edition. World Health Organization, Geneva, 2020. https://apps.who.int/iris/handle/10665/42344.
11. Copur MS and Manapuram S: Multiple primary tumors over a lifetime. Oncology (Williston Park) 33: 629384, 2019.
12. Shaitelman SF, Grills IS, Kestin LL, Ye H, Nandalur S, Huang J and Vicini FA: Rates of second malignancies after definitive local treatment for ductal carcinoma in situ of the breast. Int J Radiat Oncol Biol Phys 81: 1244-1251, 2011.
13. Goodman ZD: Neoplasms of the liver. Mod Pathol 20 (Suppl 1): S49-S60, 2007.
14. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F: Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71: 209-249, 2021.
15. Miller KD, Nogueira L, Devasia T, Mariotto AB, Yabroff KR, Jemal A, Kramer J and Siegel RL: Cancer treatment and survivorship statistics, 2022. CA Cancer J Clin 72: 409-436, 2022.
16. Xiao L, Cao T, Ou J and Liang W: Clinical characteristics and prognostic analysis of multiple primary malignant neoplasms in female patients with breast cancer or genitalia malignancies. PeerJ 10: e13528, 2022.
17. Luciani A, Ascione G, Marussi D, Oldani S, Caldiera S, Bozzoni S, Codecà C, Zonato S, Ferrari D and Foa P: Clinical analysis of multiple primary malignancies in the elderly. Med Oncol 26: 27-31, 2009.
18. Tripodi D, Cannistra' C, Gagliardi F, Casella G, Lauro A, De Luca A, Amabile MI, Palumbo P, Pironi D, Mascagni D, *et al*: Coincidental or Causal? concurrence of colorectal carcinoma with primary breast cancer. Dig Dis Sci 67: 437-444, 2022.
19. Liu Y and Cao X: Characteristics and significance of the Pre-metastatic Niche. Cancer Cell 30: 668-681, 2016.
20. Noh SK, Yoon JY, Ryoo UN, Choi CH, Sung CO, Kim TJ, Bae DS and Kim BG: A case report of quadruple cancer in a single patient including the breast, rectum, ovary, and endometrium. J Gynecol Oncol 19: 265-269, 2008.
21. National Comprehensive Cancer Network (NCCN): Breast cancer (version 2.2022). https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf.
22. Lu Y, Segelman J, Nordgren A, Lindström L, Frisell J and Martling A: Increased risk of colorectal cancer in patients diagnosed with breast cancer in women. Cancer Epidemiol 41: 57-62, 2016.
23. Phelan CM, Iqbal J, Lynch HT, Lubinski J, Gronwald J, Moller P, Ghadirian P, Foulkes WD, Armel S, Eisen A, *et al*: Incidence of colorectal cancer in BRCA1 and BRCA2 mutation carriers: Results from a follow-up study. Br J Cancer 110: 530-534, 2014.
24. Yousefi M, Nosrati R, Salmaninejad A, Dehghani S, Shahryari A and Saberi A: Organ-specific metastasis of breast cancer: Molecular and cellular mechanisms underlying lung metastasis. Cell Oncol (Dordr) 41: 123-140, 2018.
25. Babacan NA, Aksoy S, Cetin B, Ozdemir NY, Benekli M, Uyeturk U, Ali Kaplan M, Kos T, Karaca H, Oksuzoglu B, *et al*: Multiple primary malignant neoplasms: Multi-center results from Turkey. J BUON 17: 770-775, 2012.
26. Lv M, Zhang X, Shen Y, Wang F, Yang J, Wang B, Chen Z, Li P, Zhang X, Li S and Yang J: Clinical analysis and prognosis of synchronous and metachronous multiple primary malignant tumors. Medicine (Baltimore) 96: e6799, 2017.
27. Zhai C, Cai Y, Lou F, Liu Z, Xie J, Zhou X, Wang Z, Fang Y, Pan H and Han W: Multiple primary malignant tumors-A clinical analysis of 15,321 patients with malignancies at a single center in China. J Cancer 9: 2795-2801, 2018.
28. Contrera KJ, Yong V, Reddy CA, Berber E and Lorenz RR: Second primary tumors in patients with a head and neck paraganglioma. Head Neck 41: 3356-3361, 2019.