

## Risk factors and prediction of second primary cancer in primary female non-metastatic breast cancer survivors

Xiwen Qian<sup>1,\*</sup>, Huixun Jia<sup>2,\*</sup>, Yue Zhang<sup>1</sup>, Bingqing Ma<sup>1</sup>, Guoyou Qin<sup>1</sup>, Zhenyu Wu<sup>1</sup>

<sup>1</sup>Department of Biostatistics, School of Public Health, Key Laboratory of Public Health Safety and Collaborative Innovation Center of Social Risks Governance in Health, Fudan University, Shanghai, China

<sup>2</sup>Clinical Research Center, Shanghai General Hospital, Shanghai JiaoTong University School of Medicine, Shanghai, China

\*Equal contribution

**Correspondence to:** Guoyou Qin, Zhenyu Wu; **email:** [gyqin@fudan.edu.cn](mailto:gyqin@fudan.edu.cn), [zyw@fudan.edu.cn](mailto:zyw@fudan.edu.cn)

**Keywords:** breast cancer, second primary cancer, competing risk, nomogram

**Received:** June 10, 2020

**Accepted:** August 1, 2020

**Published:** October 13, 2020

**Copyright:** © 2020 Qian et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

This study aimed to investigate the risk factors of second primary cancer among female breast cancer (BC) survivors, with emphasis on the prediction of the individual risk conditioned on the patient's characteristics. We identified 208,474 BC patients diagnosed between 2004 and 2010 from the Surveillance, Epidemiology and End Results (SEER) database. Subdistribution proportional hazard model and competing-risk nomogram were used to explore the risk factors of second primary BC and non-BC, and to predict the 5- and 10-year probabilities of second primary BC. Model performance was evaluated via calibration curves and decision curve analysis. The overall 3-, 5-, and 10-year cumulative incidences for second primary BC were 0.9%, 1.6% and 4.4%, and for second primary non-BC were 2.3%, 3.9%, and 7.8%, respectively. Age over 70 years at diagnosis, black race, tumor size over 2 cm, negative hormone receptor, mixed histology, localized tumor, lumpectomy alone, and surgeries plus radiotherapy were significantly associated with increased risk of second BC. The risk of second non-BC was only related to age, race and tumor size. The proposed risk model as well as its nomogram was clinically beneficial to identify patients at high risk of developing second primary breast cancer.

### INTRODUCTION

Female breast cancer (BC) is the most common cancer in women. In the US, BC showed the highest incidence among female cancers regardless of race or ethnicity, and the mortality was second only to lung and bronchus cancer between 2012 and 2016 [1]. The American Cancer Society reported that the incidence of female BC increased slightly from 2006 to 2015 by 0.4% annually, and more than 3.5 million women had a history of invasive breast cancer by 2016 [2]. By contrast, the mortality of BC has dropped significantly by 40% from 1989 to 2016 [3], and this is attributed to improvements in early detection, diagnosis, and treatments.

With the increasing number of BC survivors, long-term negative outcomes, such as developing a second primary cancer (SPC), have become a new public health concern. Evidence shows that BC survivors have a higher risk for subsequent cancers than the general population [4–7], particularly during the first 10 years after initial diagnosis [4]. Age [8], race [9], hormone receptor status [10], and lifestyle have been reported to be related to the development of SPC among female BC survivors.

This study aimed to investigate the risk factors of developing second primary BC and non-BC among BC survivors, with emphasis on prediction of the individual probabilities of SPCs conditioned on the patient's

characteristics. In most previous studies, the risk of SPCs in BC survivors was either calculated with standard incidence ratios (SIRs) [6, 11–13], or with hazard ratio in traditional survival analysis [10, 13, 14]. The former is a typical univariate analysis, and focused on the overall risk to the general population, while the latter neglects the impact of death, which precludes the occurrence of SPCs in BC survivors, leading to an upward bias. In our study, the multivariate competing-risk model was used, in which death was considered as a competing event. Furthermore, the competing-risk model-based nomogram was provided as the tool for predicting SPC risks in female BC survivors for clinical convenience.

## RESULTS

The characteristics of the study population are listed in Table 1. Of the 208,474 FBC patients identified, 6,242 (3.0%) developed second primary BC, and 12,350 (5.9%) developed second primary non-BC. There were almost 80 sites developing SPCs in addition to breast. The top-10 most commonly diagnosed tumor sites of second non-BC were lung and bronchus, corpus uteri, melanoma of the skin, thyroid, pancreas, ovary, kidney, urinary bladder, NHL-Nodal and cecum (Supplementary Figure 1). The median and maximum follow-up time were 7.4 and 12 years, respectively.

CIFs of second primary BC and second primary non-BC are shown in Figures 1 and 2. Regarding death as a competing event, the 3-, 5- and 10-year cumulative incidences of second BC were 0.9%, 1.6% and 4.4%, respectively; the 3-, 5- and 10-year CIFs of second non-BC were 2.3%, 3.9%, and 7.8%, respectively. Female BC patients diagnosed at older age had higher incidence rate of developing SPCs ( $P < 0.001$  for both types of SPCs). Race, nodal status, tumor grade, tumor stage, and treatment were also significantly associated with cumulative incidences of both second primary BC and non-BC.

Results for the selected variables in Fine-Gray models are listed in Table 2. Compared with young patients (age < 50), patients aged over 70 years had 47% (sHR=1.47, 95%CI: 1.36 - 1.58) excessive risk of second BC, while the risk of developing second non-BC in patients aged between 50-59, 60-69 and over 70 years were 1.57 times (sHR = 1.59, 95%CI: 1.48 - 1.67), 2.56 times (sHR = 2.59, 95%CI: 2.42 - 2.72) and 4.65 times (sHR = 4.70, 95%CI: 4.39 - 4.93) higher, respectively. Black women had the highest risk (sHR=1.47, 95%CI: 1.36 - 1.58) for second BC compared with white women and Asian/Pacific Islander, and Asian or Pacific Islanders were less likely to develop second non-BC (sHR=0.84, 95%CI: 0.78 - 0.91). Tumor size over 2 cm

(sHR = 1.06, 95%CI = 1.01 - 1.12), negative HR (sHR=1.40, 95%CI: 1.32 - 1.49), mixed histology (sHR = 1.16, 95%CI = 1.07 - 1.26), lumpectomy alone (sHR=1.67, 95%CI: 1.55 - 1.81), and surgeries combined with radiotherapy were significantly associated with increased risk of second BC. Tumor size over 2 cm (sHR = 1.09, 95%CI = 1.06 - 1.14) was also related to the risk of second non-BC.

Competing-risk nomograms for based on the Fine-Gray model were constructed to predict the 5- and 10-year probabilities of second primary BC (Figure 3). Each selected variable is listed separately in the figure, with a corresponding point assigned to a given magnitude of the variable according to the point scale at the top of image. Then for every patient, the total points by summing up points of all variables were matched to the total points scale, which correspond to the 5- and 10-year predictive probabilities at the bottom of image. For example, a 65-year-old black woman was diagnosed with small (< 2 cm), HR-negative, lobular, and localized breast tumor and treated with lumpectomy alone, then she would have 3.8% and 10.4% probabilities of developing a second primary BC after initial diagnosis at 5 years and 10 years, respectively. Calibration curves for 5- and 10-year probabilities of second primary BC indicated good calibration with high correlation between predicted probabilities and observed probability of second BC (Figure 4). The c-index was 0.61 with a moderate discrimination power. DCA was applied to evaluate clinical usefulness of the prediction nomogram. Figure 5 compares the net benefit of the predictive model to those in two hypothetical scenarios: screening all BC survivors and screening none. Within a wide range, the clinical net benefit of the risk model used in 5- and 10-year second BC predictions was larger than that under the assumption of not using the model or screening all patients (Figure 5A and 5B).

## DISCUSSION

To our knowledge, the current study is the largest one to predict the risk of second primary BC and compare the risk factors for second BC and non-BC in female BC survivors. According to our population-based findings, the overall 3-, 5-, and 10-year cumulative incidences for second primary BC were 0.9%, 1.6% and 4.4% respectively, and for second primary non-BC were 2.3%, 3.9%, and 7.8%, respectively. Competing-risk events occur frequently in oncology studies. For example, 5.1% of patients died before developing an SPC in this study. Given that censoring these patients will lead to an overestimated probability of SPCs, the subdistribution competing-risk model was used. For clinical convenience, a competing-risk nomogram was proposed, and the clinical 5- or 10-year risk of

**Table 1. Baseline characteristics of study population (N=208,474).**

Variable	Censored	Stratified events, No. (%)			P
		Second BC	Second non-BC	Death	
Total	179185 (86.0)	6242 (3.0)	12350 (5.9)	10697 (5.1)	
<b>Age at diagnosis</b>					< 0.001
<50	54470 (30.4)	1678 (26.9)	1617 (13.1)	792 (7.4)	
50~	53303 (29.7)	1612 (25.8)	2543 (20.6)	1444 (13.5)	
60~	45497 (25.4)	1532 (24.5)	3696 (29.9)	2940 (27.5)	
70~	25915 (14.5)	1420 (22.7)	4494 (36.4)	5521 (51.6)	
<b>Race</b>					< 0.001
White	145755 (81.3)	4884 (78.2)	10433 (84.5)	8773 (82.0)	
Black	17980 (10.0)	865 (13.9)	1142 (9.2)	1399 (13.1)	
Asian/Pacific Islander	15450 (8.6)	493 (7.9)	775 (6.3)	525 (4.9)	
<b>Laterality</b>					0.275
Left	90873 (50.7)	3096 (49.6)	6271 (50.8)	5458 (51.0)	
Right	88312 (49.3)	3146 (50.4)	6079 (49.2)	5239 (49.0)	
<b>Tumor size (cm)</b>					< 0.001
< 2	104210 (58.2)	3711 (59.5)	7474 (60.5)	5807 (54.3)	
2 ~	74975 (41.8)	2531 (40.5)	4876 (39.5)	4890 (45.7)	
<b>Nodal status</b>					< 0.001
Negative	119287 (66.6)	4409 (70.6)	8719 (70.6)	7081 (66.2)	
Positive	59898 (33.4)	1833 (29.4)	3631 (29.4)	3616 (33.8)	
<b>HR status</b>					
Negative	35576 (19.9)	1524 (24.4)	2239 (18.1)	1992 (18.6)	
Positive	143609 (80.1)	4718 (75.6)	10111 (81.9)	8705 (81.4)	
<b>Histology</b>					< 0.001
Ductal	138225 (77.1)	4699 (75.3)	9284 (75.2)	8017 (74.9)	
Lobular	12097 (6.8)	410 (6.6)	877 (7.1)	802 (7.5)	
Mixed	18884 (10.5)	736 (11.8)	1403 (11.4)	1131 (10.6)	
Other	9979 (5.6)	397 (6.4)	786 (6.4)	747 (7.0)	
<b>Grade</b>					< 0.001
I	38483 (21.5)	1352 (21.7)	2969 (24.0)	2410 (22.5)	
II	74945 (41.8)	2562 (41.0)	5311 (43.0)	4723 (44.2)	
III	64083 (35.8)	2271 (36.4)	3970 (32.1)	3464 (32.4)	
IV	1674 (0.9)	57 (0.9)	100 (0.8)	100 (0.9)	
<b>Tumor stage</b>					< 0.001
Localized	116250 (64.9)	4307 (69.0)	8506 (68.9)	6840 (63.9)	
Regional	62935 (35.1)	1935 (31.0)	3844 (31.1)	3857 (36.1)	
<b>Treatment</b>					< 0.001
Mastectomy alone	53094 (29.6)	1431 (22.9)	3584 (29.0)	4024 (37.6)	
Mastectomy +Radiotherapy	18898 (10.5)	565 (9.1)	1082 (8.8)	918 (8.6)	
Lumpectomy alone	22661 (12.6)	1056 (16.9)	1734 (14.0)	1666 (15.6)	
Lumpectomy +Radiotherapy	84532 (47.2)	3190 (51.1)	5950 (48.2)	4089 (38.2)	

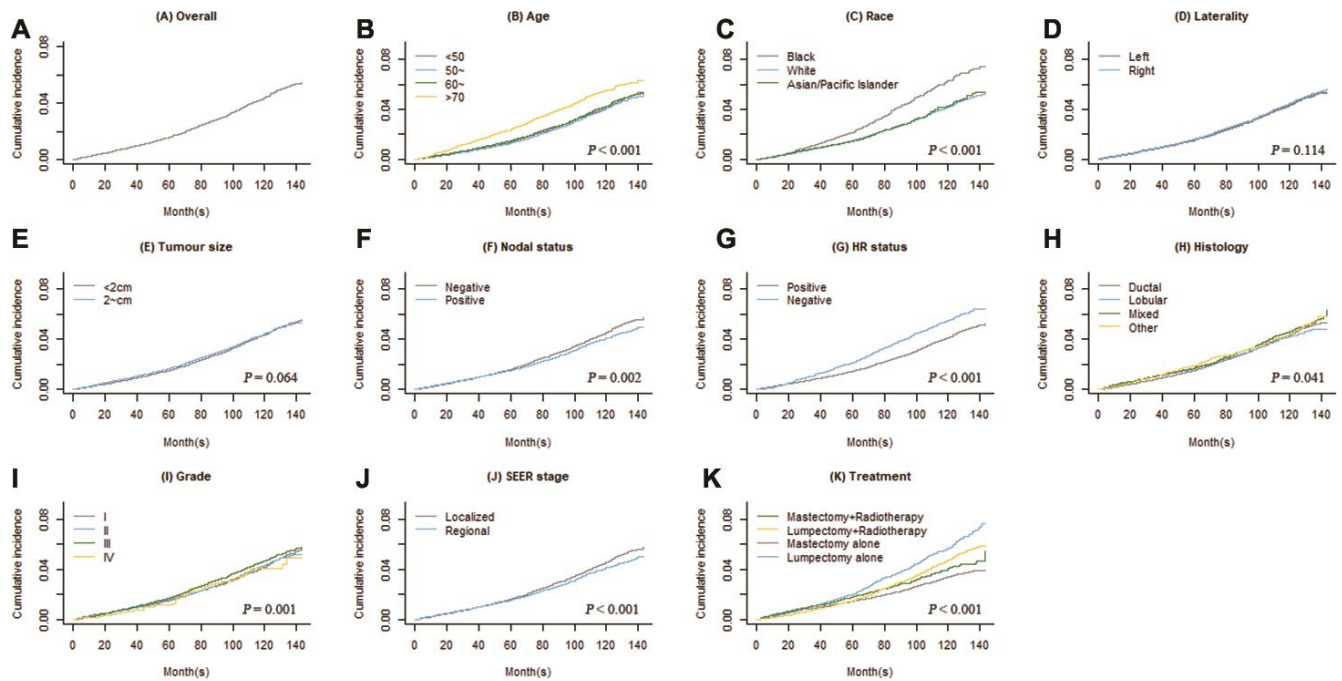
Abbreviation: BC, breast cancer; HR, hormone receptor.

developing second BC could be estimated. Identifying subpopulations and individuals with distinct risk of SPCs using the proposed risk model can greatly benefit patients through decision making, because a patient with high risk of SPCs might be asked to consider more subsequent screening, changing adverse living habits or choosing optimal treatment for the current BC to avoid increasing risk from the side

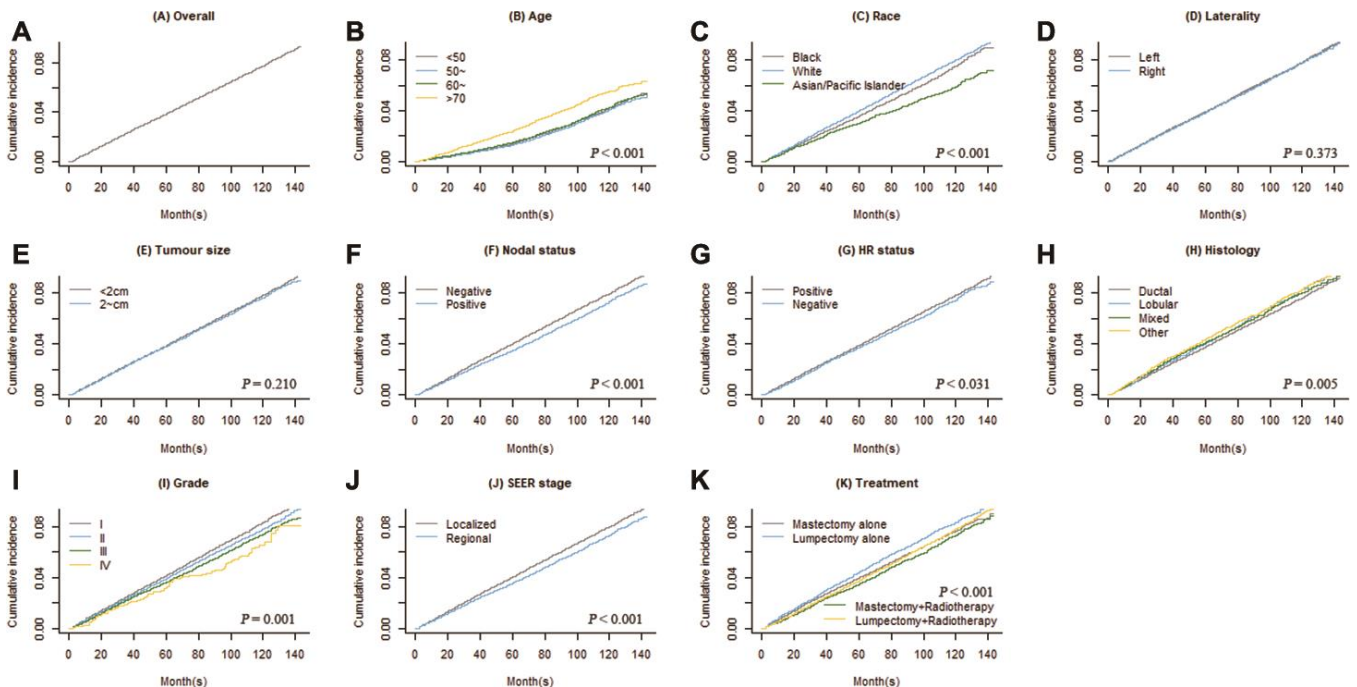
effect. Ideally, for those with high risk of developing SPCs, active and regular surveillance could benefit them from early prevention, early detection and early treatment, to improve their quality of life. High-risk BC patients would be identified according to our proposed nomogram, and only for those with high-risk patients, active strategy could be considered. In practice, the implementation of follow-up surveillance

strategy not only depends on the risk of subsequent SPC, but also on other factors, such as patient's financial arrangement, medical insurance and access to healthcare.

Results from the multivariate competing-risk models indicated that age of initial diagnosis and race were significantly related to the risks of both second primary BC and non-BC. Risk differences of age at diagnosis



**Figure 1.** Overall cumulative incidence function (CIF) curve of second primary breast cancer (BC) (A) and CIF curves grouped by each covariate (B–K).



**Figure 2.** Overall cumulative incidence function (CIF) curve of second primary non-BC (A) and CIF curves grouped by each covariate (B–K).

**Table 2. Factors associated with the risk of second primary breast cancer (BC) and non-BC among BC patients in final predictive models.**

	sHR	95% CI	P
<b>Second primary BC</b>			
<b>Age at diagnosis</b>			
< 50	Ref		
50 ~	0.95	0.89 - 1.02	0.143
60 ~	1.04	0.97 - 1.12	0.267
70 ~	1.47	1.36 - 1.58	< 0.001
<b>Race</b>			
White	Ref		
Black	1.47	1.36 - 1.58	< 0.001
Asian/Pacific Islander	1.06	0.98 - 1.17	0.183
<b>Tumor size</b>			
<2 cm	Ref		
2~cm	1.06	1.01 - 1.12	0.030
<b>HR status</b>			
Positive	Ref		
Negative	1.40	1.32 - 1.49	< 0.001
<b>Histology</b>			
Ductal	Ref		
Lobular	1.09	0.98 - 1.21	0.107
Mixed	1.16	1.07 - 1.26	< 0.001
Other	1.03	0.93 - 1.14	0.590
<b>Tumor stage</b>			
Localized	Ref		
Regional	0.93	0.87 - 0.99	0.048
<b>Treatment</b>			
Mastectomy alone	Ref		
Mastectomy + Radiotherapy	1.27	1.15 - 1.41	< 0.001
Lumpectomy + Radiotherapy	1.37	1.28 - 1.46	< 0.001
Lumpectomy alone	1.67	1.55 - 1.81	< 0.001
<b>Second Primary non-BC</b>			
<b>Age at diagnosis</b>			
< 50	Ref		
50 ~	1.57	1.48 - 1.67	< 0.001
60 ~	2.56	2.42 - 2.72	< 0.001
70 ~	4.65	4.39 - 4.93	< 0.001
<b>Race</b>			
White	Ref		
Black	1.01	0.95 - 1.08	0.716
Asian/Pacific Islander	0.84	0.78 - 0.91	< 0.001
<b>Tumor size (cm)</b>			
< 2	Ref		
2 ~	1.09	1.06 - 1.14	< 0.001

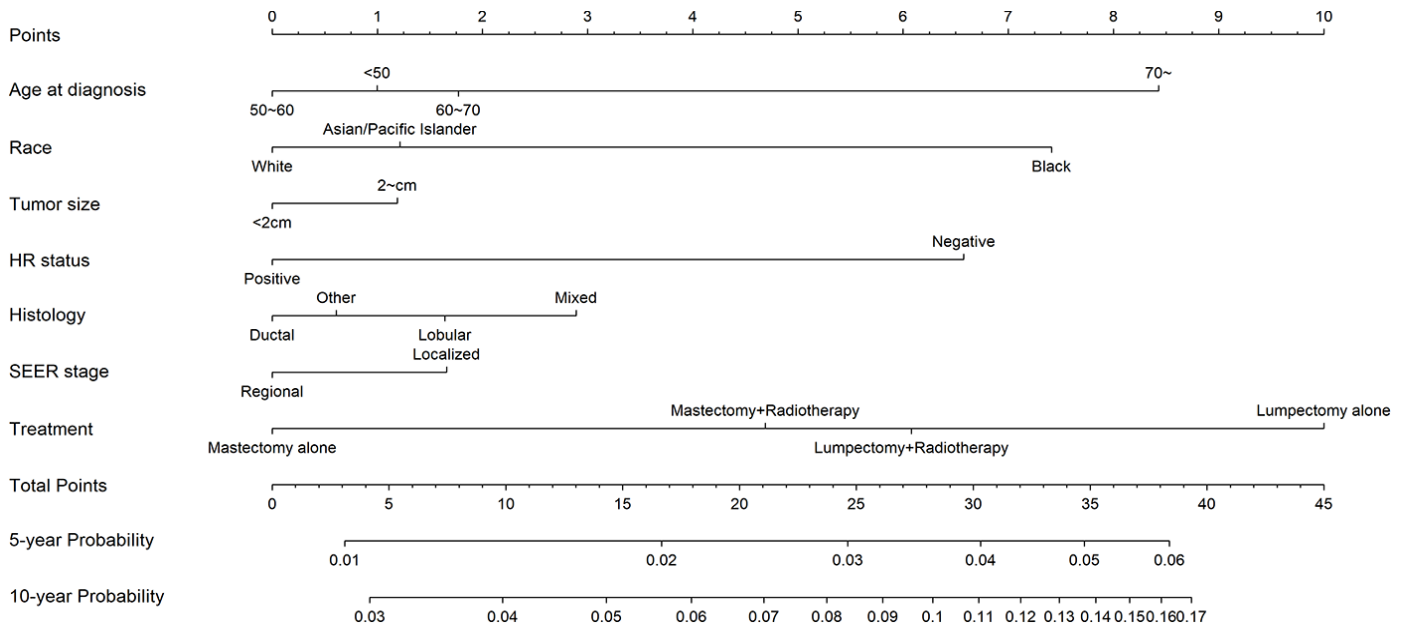
Abbreviation: BC, breast cancer; sHR, hazard ratio from subdistribution model; HR, hormone receptor.

have been extensively examined but remains debatable. Some studies showed that young patients (< 50 years) had greater predisposition to SPCs than the old patients (≥ 50 years) [4, 6, 11, 12, 15]. However, these findings

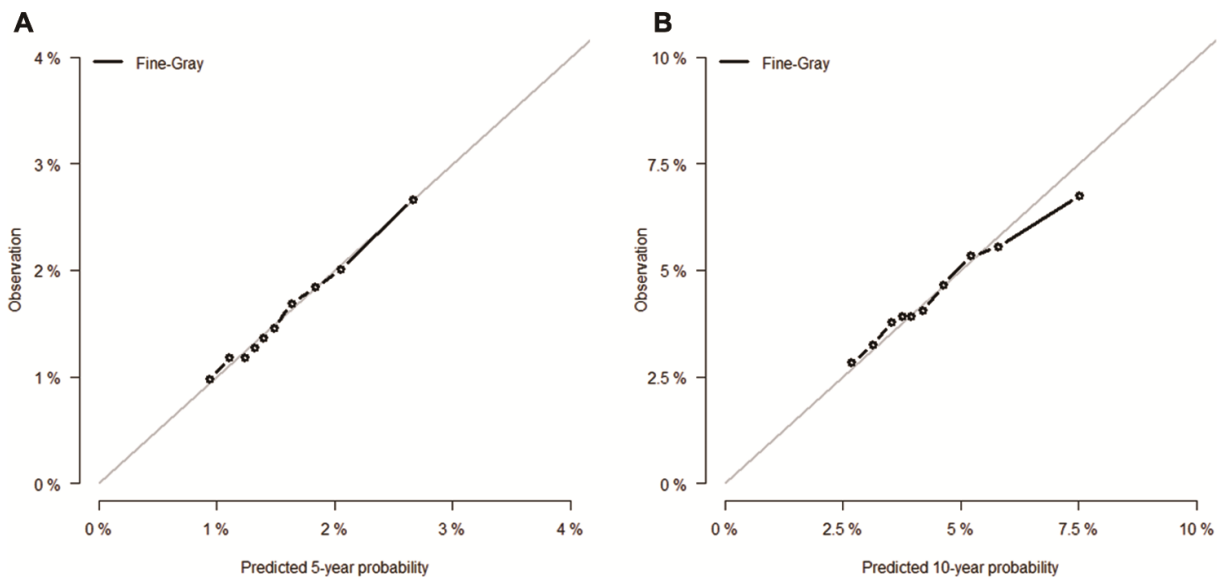
concluded from comparisons between every age group of general population and of the study population without considering other factors simultaneously, thereby SIR cannot comprehensively measure the effect

of age. Instead, multivariate Cox proportional hazard regression or its extensions would be more appropriate to identify risk differences based on multiple factors. Our research found that an older age was associated with higher risk of second primary BC and non-BC in BC patients, which is consistent with other multivariate studies [8, 10, 13, 14]. It is well documented that

advanced age is among the most important risk factors for many age specific cancer types due to the functional decline in immune system, increased susceptibility to carcinogen, reduced DNA repair, and long-term effect of harmful lifestyle. In addition, that older patients tended to develop more SPCs may also be affected by less systemic therapy due to their poor physical conditions.



**Figure 3.** Competing-risk nomogram for predicting 5- and 10-year risks of second primary breast cancer (BC) in female BC survivors.



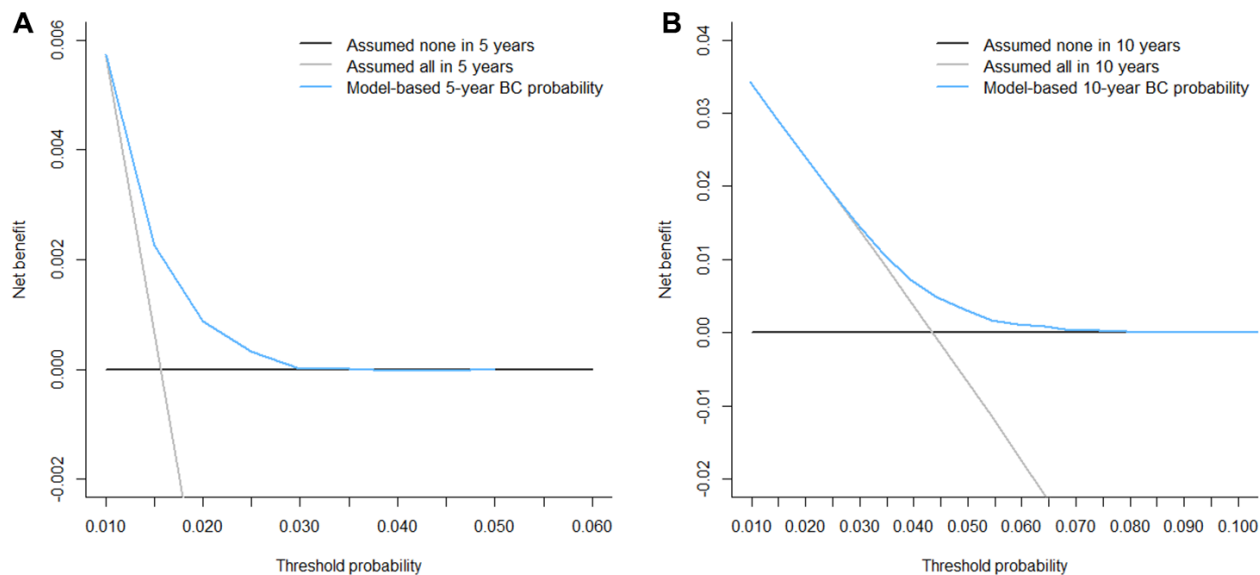
**Figure 4.** Calibration curves for (A) 5- and (B) 10-year predictions from Fine-Gray model. X-axes indicate predicted 5- or 10-year probabilities; Y-axes indicate actual observations.

Systemic therapy, including chemotherapy, hormonal therapy and immunotherapy, works throughout the entire body for treating or preventing metastasis, and it may also inhibit the development of new tumors before their diagnosis.

Race is also an important demographic predictor for SPC risk. The incidence and risk of second BC was higher among black women than white women and Asian/Pacific Islanders. Other studies also confirmed this racial disparity in the risk of second primary BC [9, 16, 17], which might be explained by the combination of genetic or biologic, clinical and socioeconomic factors. For example, some genetic differences in *BRCA1*, *BRCA2* and *p53* mutations between African Americans and Caucasians in the US have been shown to be associated with different pathological characteristics and tumor prognosis [18]. Moreover, a lower socioeconomic status in any race tends to cause worse tumor pathological characteristics, inappropriate treatments and less healthcare utilization before and after first primary BC [19].

We also observed increased second primary BC risk in BC subtypes classified by clinical tumor characteristics, such as tumor size, HR status, tumor histology and SEER stage. HR-negative BC patients had a higher risk of SPC, which is consistent with previous studies [10, 20, 21]. However, it remains unclear whether the correlation between HR status of the initial BC and risk of second BC reflects an

underlying genetic predisposition or exposure to hormone or other risk factors. A review concluded that *BRCA1* mutation carriers were more likely to have triple-negative (ER-, PR- and HER2-negative) BC [22], and these women also had increased risk of developing SPC [23]. In addition, previous studies showed that the distribution of HR status in the female BC patients might be caused by racial or ethnic disparities [24, 25]. Specifically, triple-negative patients were more likely to be non-Hispanic black and Hispanic, while white women had the highest incidence of HR-positive BC subtype. As we discussed previously, black women suffered higher risk of second primary BC than white and Asian or Pacific Islanders; hence, the interaction between HR status and race may partly explain the higher risk of second BC. In summary, it is important to emphasize that in addition to the poor prognosis of HR-negative BC patients, they may also have an elevated risk of developing second primary BC, which would in turn lead to a poor survival. Other studies also examined the association between human epidermal growth factor receptor 2 (HER2) status and the risk of SPC which still remains further research. One previous study reported that those with HER2-overexpressing (ER-/HER2+) and triple-negative BC had elevated risks of developing second primary contralateral breast cancer [21]. A recent study found those with HER2-BC and without adjuvant trastuzumab therapy had higher risk of contralateral breast cancer [26]. While another study claimed that HER2 status did not seem to be a marker of risk for a SPC [27]. For second primary



**Figure 5. Decision curve analysis for the risk models for second primary breast cancer (BC).** The decision curves shows that within the threshold probabilities ((A) 1.0% - 3.0% for 5-year and (B) 2.8% - 7.5% for 10-year prediction of second primary BC, respectively), using the competing-risk model to predict the probability of developing second primary BC can produce more benefit than treating either all or no patient would have second BC.

non-BC, it was reported that positive HER2 status increased the risk of digestive system and thyroid tumors [28]. Other pathological features, such as large tumor size, reflected an advanced tumor stage. BC patients diagnosed with advanced tumor might have lived for a longer survival time after tumor cell appeared than patients diagnosed at early stage, so they could have higher risk of recurrence or SPCs. Besides, the misclassification of SPC and metastasis also might happen.

In the current study, the effect of surgical type and radiation treatment were investigated. Lumpectomy plus radiotherapy is one of the common treatment strategies and is likely to be as effective as mastectomy. Although radiotherapy was regarded as an important risk factor for SPCs, it showed no significant effect on non-BC in the final model and less adverse effect than lumpectomy alone on developing second BC. It was reported by other studies that radiotherapy combined with surgery did not increase the risk of second primary BC or non-BC compared to surgery-alone treatment [29], and that radiotherapy after surgery reduced the long-term risk of recurrence [30]. What is more, the progress in technology can provide better radiation quality and dose conformality to target, hence better preserve the adjacent normal tissue and reduce the risk of radiotherapy-related SPCs like lung cancer and leukemia. The decreasing adverse effect of radiotherapy has been reported in several previous studies [31, 32]. Patients treated with lumpectomy alone had the highest risk of second BC, which may imply the drawback of this surgery type that BC would come back if tumor cells were not removed completely.

The current study had strengths with respect to its methodology and model ability of identifying high-risk SPC patients. Limitations mainly came from the absence of some important variables in the SEER database, such as detailed treatment information including chemotherapy, radicality of breast-conserving surgery and so on, and factors that influence cancer incidence including smoking status, sociodemographic status and genetic factors. A moderate C-index may also result from the absence of detailed information. Besides, the quality of collected data may differ among registry sites over time. Future work will focus on external validation with other populations and perfecting our proposed model by including more covariates.

## CONCLUSIONS

Important risk factors for second primary BC after initial primary BC diagnosis included age at diagnosis, black race, large tumor size, negative HR status, mixed tumor histology, localized tumor and receiving

lumpectomy and radiotherapy. While the risk of second primary non-BC was less relevant to initial breast tumor characteristics and treatment. The proposed nomogram was clinically beneficial to identify patients at high risk of developing SPCs, so they may benefit from further efforts targeted at improving access to better healthcare and SPC prevention.

## MATERIALS AND METHODS

### Data source and study population

Data were collected from the Surveillance, Epidemiology, and End Results (SEER) Program (2018 submission). The SEER program collects information on demographics, tumor characteristics, and treatments of patients from various regions in the US since 1973 [33]. Female BC patients were included in the study if they were (1) diagnosed as primary BC, (2) aged between 20 and 80 years, (3) diagnosed before 2011 (to ensure a minimum of 5-year follow-up after initial diagnosis), and (4) diagnosed after 2003 because variables like estrogen receptor (ER) and progesterone receptor (PR) were recorded since 2004. Patients were subsequently excluded if (1) information was obtained from death certificate or autopsy, (2) demographic or clinical information was missing or unknown, (3) treatment information were not recorded or unspecific, (4) distant metastases occurred, and (5) diagnosed with bilateral tumors. A total of 208,474 female non-metastatic BC patients diagnosed between 2004 and 2010 and who were followed up over 5 years were identified as the final study population.

### Definition of SPC

SPC was defined according to the SEER guidelines, which takes into account the tumor site, tumor behavior (in situ or invasive), histology, date of diagnosis, and laterality of paired organs [34]. The SEER rules mainly comprise the followings:

- 1) A new lesion with same histologic type as the previous tumor in the same site and occurring synchronously (within 2 months) is to be regarded as a single tumor, not a new primary one.
- 2) A new lesion with different histologic type as the previous tumor in the same site and occurring synchronously (within 2 months) is to be regarded as a new primary tumor.
- 3) A new lesion in the same site occurring metachronously (2 months or more after the primary diagnosis) is counted as a new primary tumor regardless of histologic type, unless it is confirmed as a metastatic lesion.



4) A new lesion in different site with either the same or different histologic type is to be considered as a new primary tumor regardless of time, unless it is stated to be a metastatic lesion.

5) For paired organs, synchronous bilateral tumor with only one histologic type is considered as a single primary tumor; bilateral tumor with two different histologic types is considered two primaries unless stated to the contrary.

## Outcome and covariates

The primary outcome of the current study was the occurrence of SPC in female BC patients, and SPCs were further classified as second primary BC and second primary non-BC. Death as the result of any cause before SPC was defined as a competing event. Demographic covariates included age at initial diagnosis of BC and race (White, Black and Asian/Pacific Islander). We categorize age into <50, 50~, 60~, and 70~, because 1) female breast cancer patients after menopause are at higher risk of developing a new malignancy [14], 2) menopause is reported starts from the age of 50 in the United States [35], 3) 10-year-old age groups are commonly used in related studies [14, 28, 36]. Tumor characteristic covariates referred to laterality, tumor size, nodal status, hormone receptor (HR) status (Positive: either ER or PR was positive; Negative: both ER and PR were negative), histology (Ductal, Lobular, Mixed and Other), grade (I: well-differentiated, II: moderately-differentiated, III: poorly-differentiated and IV: undifferentiated), SEER stage (Localized and Regional). In this study population, radiotherapy was given to 78.1% of patients treated with lumpectomy, and 72.9% of patients treated with mastectomy were not given radiotherapy. Therefore, treatment was categorized into four types: mastectomy alone, mastectomy plus radiotherapy, lumpectomy alone and lumpectomy plus radiotherapy.

## Statistical analyses

Possible differences of patients' characteristics between event groups were examined by Chi-square test. The overall cumulative incidence function (CIF) of developing an SPC as well as CIFs grouped by each covariate was calculated. We plotted the CIF curves and performed the Gray's test to identify differences between groups.

### *Competing-risk modelling and nomogram*

The Fine-Gray subdistribution proportional hazard model [37], an extension of conventional Cox proportional hazard model, is commonly used in the presence of competing risk and is suggested to be

appropriate for quantifying the effect of covariates on the incidence of the outcome and predicting incidence over time [38, 39]. Therefore, in this study, the subdistribution proportional hazard model was applied to explore risk factors and to predict the unbiased risk probability of developing an SPC. Variable selection was determined via stepwise backward elimination method with Akaike information criteria (AIC). Subdistribution hazard ratio (sHR) was used to assess the effect of each covariate on the risk of SPCs. A competing-risk nomogram was subsequently built to predict the 5- and 10-year probabilities of developing SPCs. Each covariate was assigned with a score to indicate the variable importance, and by summing up all scores of selected variables, the predictive probabilities of developing an SPC for a certain BC patient can be obtained.

### *Model performance and evaluation*

For validation, we used calibration with a bootstrap cross-validation method to evaluate the model performance. The predictive model was trained on 200 bootstrap samples drawn with replacement of the same size as the original data. The model was assessed in the observations that are not in the bootstrap sample. The concordance index (C-index) was calculated to measure the ability of the model to distinguish outcomes. Furthermore, decision curve analysis (DCA) [40], which is a novel method of evaluating diagnostic and prognostic prediction models, was applied to assess the clinical usefulness and magnitude of benefit of the proposed competing-risk model. A model is clinically valuable in a certain range of threshold probabilities, if it produces a larger net benefit through identifying patients with high risk of SPC than assuming all or none of the patients will develop an SPC.

All the statistical analyses and graph plotting were performed in R 3.5.1 (<http://www.r-project.org>).

## Abbreviations

BC: female breast cancer; SPC: second primary cancer; SIR: standard incidence ratio; SEER: Surveillance: Epidemiology: and End Results Program; HR: hormone receptor; ER: estrogen receptor; PR: progesterone receptor; CIF: cumulative incidence function; sHR: subdistribution hazard ratio; DCA: decision curve analysis.

## AUTHOR CONTRIBUTIONS

Xiwen Qian, Huixun Jia, Guoyou Qin and Zhenyu Wu conceived the idea; Xiwen Qian, Huixun Jia, Yue Zhang, Bingqing Ma, Guoyou Qin and Zhenyu Wu performed the statistical analyses and interpreted the

results; Xiwen Qian, Huixun Jia, Guoyou Qin and Zhenyu Wu participated in the writing and editing of the manuscript. All authors approved the final manuscript.

## CONFLICTS OF INTEREST

Xiwen Qian, Huixun Jia, Yue Zhang, Bingqing Ma, Guoyou Qin and Zhenyu Wu declare that he has no conflict of interest.

This article does not contain any studies with human or animal subjects.

## FUNDING

This study was funded by the Scientific Project of Shanghai Municipal Health Commission (201940151) and the National Natural Science Foundation of China (11871164).

## REFERENCES

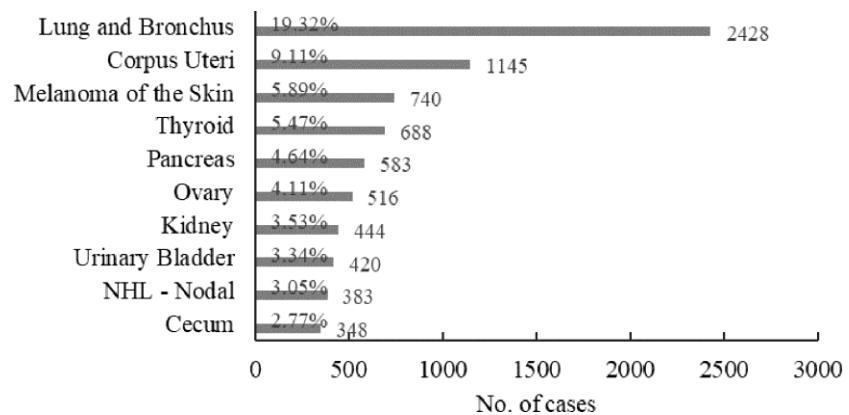
1. Howlader N, Noone A, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis D, Chen H, Feuer E, Cronin Ke. SEER Cancer Statistics Review, 1975-2016. National Cancer Institute [https://seer.cancer.gov/archive/csr/1975\\_2016/](https://seer.cancer.gov/archive/csr/1975_2016/). 2019.
2. Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, Stein KD, Alteri R, Jemal A. Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin*. 2016; 66:271–89. <https://doi.org/10.3322/caac.21349> PMID:27253694
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019; 69:7–34. <https://doi.org/10.3322/caac.21551> PMID:30620402
4. Molina-Montes E, Requena M, Sánchez-Cantalejo E, Fernández MF, Arroyo-Morales M, Espín J, Arrebola JP, Sánchez MJ. Risk of second cancers cancer after a first primary breast cancer: a systematic review and meta-analysis. *Gynecol Oncol*. 2015; 136:158–71. <https://doi.org/10.1016/j.ygyno.2014.10.029> PMID:25448459
5. Jung HK, Park S, Kim NW, Lee JE, Kim Z, Han SW, Hur SM, Kim SY, Lim CW, Lee MH, Lee J. Development of second primary cancer in Korean breast cancer survivors. *Ann Surg Treat Res*. 2017; 93:287–92. <https://doi.org/10.4174/ast.2017.93.6.287> PMID:29250506
6. Silverman BG, Lipshitz I, Keinan-Boker L. Second primary cancers after primary breast cancer diagnosis in Israeli women, 1992 to 2006. *J Glob Oncol*. 2016; 3:135–42. <https://doi.org/10.1200/JGO.2016.003699> PMID:28717751
7. Kamigaki Y, Kawakami K. Risk of second cancer after initial treatment of breast cancer: an osaka cancer registry database study. *Oncol Lett*. 2011; 2:963–73. <https://doi.org/10.3892/ol.2011.339> PMID:22866158
8. Ricceri F, Fasanelli F, Giraudo MT, Sieri S, Tumino R, Mattiello A, Vagliano L, Masala G, Quirós JR, Travier N, Sánchez MJ, Larranaga N, Chirlaque MD, et al. Risk of second primary Malignancies in women with breast cancer: results from the european prospective investigation into cancer and nutrition (EPIC). *Int J Cancer*. 2015; 137:940–48. <https://doi.org/10.1002/ijc.29462> PMID:25650288
9. Calip GS, Law EH, Ko NY. Racial and ethnic differences in risk of second primary cancers among breast cancer survivors. *Breast Cancer Res Treat*. 2015; 151:687–96. <https://doi.org/10.1007/s10549-015-3439-7> PMID:26012645
10. Buist DS, Abraham LA, Barlow WE, Krishnaraj A, Holdridge RC, Sickles EA, Carney PA, Kerlikowske K, Geller BM, and Breast Cancer Surveillance Consortium. Diagnosis of second breast cancer events after initial diagnosis of early stage breast cancer. *Breast Cancer Res Treat*. 2010; 124:863–73. <https://doi.org/10.1007/s10549-010-1106-6> PMID:20700648
11. Raymond JS, Hogue CJ. Multiple primary tumours in women following breast cancer, 1973-2000. *Br J Cancer*. 2006; 94:1745–50. <https://doi.org/10.1038/sj.bjc.6603172> PMID:16721370
12. Molina-Montes E, Pollán M, Payer T, Molina E, Dávila-Arias C, Sánchez MJ. Risk of second primary cancer among women with breast cancer: a population-based study in granada (Spain). *Gynecol Oncol*. 2013; 130:340–45. <https://doi.org/10.1016/j.ygyno.2013.04.057> PMID:23648471
13. Li Z, Wu Q, Song J, Zhang Y, Zhu S, Sun S. Risk of Second Primary Female Genital Malignancies in Women with Breast Cancer: a SEER Analysis. *Horm Cancer*. 2018; 9:197–204. <https://doi.org/10.1007/s12672-018-0330-0> PMID:29556910
14. Corso G, Veronesi P, Santomauro GI, Maisonneuve P, Morigi C, Peruzzotti G, Intra M, Sacchini V, Galimberti V. Multiple primary non-breast tumors in breast cancer survivors. *J Cancer Res Clin Oncol*. 2018; 144:979–86. <https://doi.org/10.1007/s00432-018-2621-9> PMID:29502170

15. Evans HS, Lewis CM, Robinson D, Bell CM, Møller H, Hodgson SV. Incidence of multiple primary cancers in a cohort of women diagnosed with breast cancer in southeast England. *Br J Cancer*. 2001; 84:435–40.  
<https://doi.org/10.1054/bjoc.2000.1603>  
PMID:11161413
16. Stark A, Stapp R, Raghunathan A, Yan X, Kirchner HL, Griggs J, Newman L, Chitale D, Dick A. Disease-free probability after the first primary ductal carcinoma in situ of the breast: a comparison between African-American and White-American women. *Breast Cancer Res Treat*. 2012; 131:561–70.  
<https://doi.org/10.1007/s10549-011-1742-5>  
PMID:21874310
17. Liu Y, Colditz GA, Gehlert S, Goodman M. Racial disparities in risk of second breast tumors after ductal carcinoma in situ. *Breast Cancer Res Treat*. 2014; 148:163–73.  
<https://doi.org/10.1007/s10549-014-3151-z>  
PMID:25261293
18. Ademuyiwa FO, Olopade OI. Racial differences in genetic factors associated with breast cancer. *Cancer Metastasis Rev*. 2003; 22:47–53.  
<https://doi.org/10.1023/a:1022259901319>  
PMID:12716036
19. Cross CK, Harris J, Recht A. Race, socioeconomic status, and breast carcinoma in the U.S: what have we learned from clinical studies. *Cancer*. 2002; 95:1988–99.  
<https://doi.org/10.1002/cncr.10830>  
PMID:12404294
20. Li D, Weng S, Zhong C, Tang X, Zhu N, Cheng Y, Xu D, Yuan Y. Risk of second primary cancers among long-term survivors of breast cancer. *Front Oncol*. 2020; 9:1426.  
<https://doi.org/10.3389/fonc.2019.01426>  
PMID:31998630
21. Saltzman BS, Malone KE, McDougall JA, Daling JR, Li CI. Estrogen receptor, progesterone receptor, and HER2-neu expression in first primary breast cancers and risk of second primary contralateral breast cancer. *Breast Cancer Res Treat*. 2012; 135:849–55.  
<https://doi.org/10.1007/s10549-012-2183-5>  
PMID:22903686
22. Chen H, Wu J, Zhang Z, Tang Y, Li X, Liu S, Cao S, Li X. Association between BRCA status and triple-negative breast cancer: a meta-analysis. *Front Pharmacol*. 2018; 9:909.  
<https://doi.org/10.3389/fphar.2018.00909>  
PMID:30186165
23. Ford D, Easton DF, Bishop DT, Narod SA, Goldgar DE. Risks of cancer in BRCA1-mutation carriers. Breast cancer linkage consortium. *Lancet*. 1994; 343:692–95.  
[https://doi.org/10.1016/S0140-6736\(94\)91578-4](https://doi.org/10.1016/S0140-6736(94)91578-4)  
PMID:7907678
24. Howlander N, Altekruse SF, Li CI, Chen VW, Clarke CA, Ries LA, Cronin KA. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *J Natl Cancer Inst*. 2014; 106:dju055.  
<https://doi.org/10.1093/jnci/dju055> PMID:24777111
25. Ooi SL, Martinez ME, Li CI. Disparities in breast cancer characteristics and outcomes by race/ethnicity. *Breast Cancer Res Treat*. 2011; 127:729–38.  
<https://doi.org/10.1007/s10549-010-1191-6>  
PMID:21076864
26. Giardiello D, Steyerberg EW, Hauptmann M, Adank MA, Akdeniz D, Blomqvist C, Bojesen SE, Bolla MK, Brinkhuis M, Chang-Claude J, Czene K, Devilee P, Dunning AM, et al. Prediction and clinical utility of a contralateral breast cancer risk model. *Breast Cancer Res*. 2019; 21:144.  
<https://doi.org/10.1186/s13058-019-1221-1>  
PMID:31847907
27. Bessonova L, Taylor TH, Mehta RS, Zell JA, Anton-Culver H. Risk of a second breast cancer associated with hormone-receptor and HER2/neu status of the first breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2011; 20:389–96.  
<https://doi.org/10.1158/1055-9965.EPI-10-1016>  
PMID:21217087
28. Marcheselli R, Marcheselli L, Cortesi L, Bari A, Cirilli C, Pozzi S, Ferri P, Napolitano M, Federico M, Sacchi S. Risk of second primary Malignancy in breast cancer survivors: a nested population-based case-control study. *J Breast Cancer*. 2015; 18:378–85.  
<https://doi.org/10.4048/jbc.2015.18.4.378>  
PMID:26770245
29. Shaitelman SF, Grills IS, Kestin LL, Ye H, Nandalur S, Huang J, Vicini FA. Rates of second Malignancies after definitive local treatment for ductal carcinoma in situ of the breast. *Int J Radiat Oncol Biol Phys*. 2011; 81:1244–51.  
<https://doi.org/10.1016/j.ijrobp.2010.07.2005>  
PMID:21030159
30. Darby S, McGale P, Correa C, Taylor C, Arriagada R, Clarke M, Cutter D, Davies C, Ewertz M, Godwin J, Gray R, Pierce L, Whelan T, et al, and Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials. *Lancet*. 2011; 378:1707–16.  
[https://doi.org/10.1016/S0140-6736\(11\)61629-2](https://doi.org/10.1016/S0140-6736(11)61629-2)  
PMID:22019144

31. Mellemkjaer L, Friis S, Olsen JH, Scélo G, Hemminki K, Tracey E, Andersen A, Brewster DH, Pukkala E, McBride ML, Kliever EV, Tonita JM, Kee-Seng C, et al. Risk of second cancer among women with breast cancer. *Int J Cancer*. 2006; 118:2285–92.  
<https://doi.org/10.1002/ijc.21651> PMID:16342146
32. Brown LM, Chen BE, Pfeiffer RM, Schairer C, Hall P, Storm H, Pukkala E, Langmark F, Kaijser M, Andersson M, Joensuu H, Fosså SD, Travis LB. Risk of second non-hematological Malignancies among 376,825 breast cancer survivors. *Breast Cancer Res Treat*. 2007; 106:439–51.  
<https://doi.org/10.1007/s10549-007-9509-8> PMID:17277968
33. Institute NC. About the SEER Program. <https://seer.cancer.gov/about/>. 2018.
34. Fritz A RLe. The SEER Program Code Manual, 3rd Edition. Bethesda, Md: National Cancer Institute. 1998; No. 98–1999.
35. Takahashi TA, Johnson KM. Menopause. *Med Clin North Am*. 2015; 99:521–34.  
<https://doi.org/10.1016/j.mcna.2015.01.006> PMID:25841598
36. Mellemkjær L, Christensen J, Frederiksen K, Pukkala E, Weiderpass E, Bray F, Friis S, Andersson M, Olsen JH. Risk of primary non-breast cancer after female breast cancer by age at diagnosis. *Cancer Epidemiol Biomarkers Prev*. 2011; 20:1784–92.  
<https://doi.org/10.1158/1055-9965.EPI-11-0009> PMID:21653644
37. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*. 1999; 94:496–509.  
<https://doi.org/10.1080/01621459.1999.10474144>
38. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol*. 2009; 170:244–56.  
<https://doi.org/10.1093/aje/kwp107> PMID:19494242
39. Austin PC, Fine JP. Practical recommendations for reporting fine-gray model analyses for competing risk data. *Stat Med*. 2017; 36:4391–400.  
<https://doi.org/10.1002/sim.7501> PMID:28913837
40. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006; 26:565–74.  
<https://doi.org/10.1177/0272989X06295361> PMID:17099194

## SUPPLEMENTARY MATERIALS

### Supplementary Figure



Supplementary Figure 1. The top-10 most commonly diagnosed sites of second primary non-BC.