

Risk assessment model for invasive breast cancer in Hong Kong women

Feng Wang, PhD^a, Juncheng Dai, PhD^{a,b}, Mengjie Li, PhD^a, Wing-cheong Chan, MD^c, Carol Chi-hei Kwok, MD^d, Siu-lan Leung, MD^e, Cherry Wu, MD^f, Wentao Li, MD^a, Wai-cho Yu, MD^g, Koon-ho Tsang, MD^h, Sze-hong Law, MDⁱ, Priscilla Ming-yi Lee, MPH^a, Carmen Ka-man Wong, MD^a, Hongbing Shen, PhD^b, Samuel Yeung-shan Wong, MD, MPH^a, Xiaohong R. Yang, PhD, MPH^j, Lap Ah Tse, PhD^{a,*}

Abstract

No risk assessment tool is available for identifying high risk population of breast cancer (BCa) in Hong Kong. A case-control study including 918 BCa cases and 923 controls was used to develop the risk assessment model among Hong Kong Chinese women.

Each participant received an in-depth interview to obtain their lifestyle and environmental risk factors. Least absolute shrinkage and selection operator (LASSO) selection model was used to select the optimal risk factors (LASSO-model). A risk score system was constructed to evaluate the cumulative effects of selected factors. Bootstrap simulation was used to test the internal validation of the model. Model performance was evaluated by receiver-operator characteristic curves and the area under the curve (AUC).

Age, number of parity, number of BCa cases in 1st-degree relatives, exposure to light at night, and sleep quality were the common risk factors for all women. Alcohol drinking was included for premenopausal women; body mass index, age at menarche, age at 1st give birth, breast feeding, using of oral contraceptive, hormone replacement treatment, and history of benign breast diseases were included for postmenopausal women. The AUCs were 0.640 (95% CI, 0.598–0.681) and 0.655 (95% CI, 0.621–0.653) for pre- and postmenopausal women, respectively. Further subgroup evaluation revealed that the model performance was better for women aged 50 to 70 years or ER-positive.

This BCa risk assessment tool in Hong Kong Chinese women based on LASSO selection is promising, which shows a slightly higher discriminative accuracy than those developed in other populations.

Abbreviations: AUC = area under the curve, BCa = breast cancer, BMI = body mass index, ER = estrogen receptor, KoBRCAT = Korean breast cancer risk assessment tool, LAN = light at night, LASSO = least absolute shrinkage and selection operator, OPT-model = optional model, ROC = receiver-operator characteristic.

Keywords: breast cancer, estrogen receptor, LASSO model, light at night, menopause, risk assessment

1. Introduction

Breast cancer (BCa) is the most common cancer among Hong Kong women. About 6% of Hong Kong women (1 in 17) will develop BCa during their lives.^[1] Although this lifetime risk is half of that (1 in 8) in the United States,^[2] the BCa incidence rate

of Hong Kong women has almost doubled during recent 3 decades.^[3] It has been reported that the BCa incidence rates of recent Chinese generations in Singapore and Taiwan (with similar economic development as Hong Kong) were even higher than those in the United States.^[4] Hence, there is an urgent need

Editor: Chun Gao.

FW and JD contributed equally to this work.

Funding/support: This work was supported by Research Grants Council of Hong Kong (Grant number 474811).

The funding source had no role in the study design, data collection, data analysis, or interpretation of the findings.

The authors have no conflicts of interest to disclose.

Supplemental Digital Content is available for this article.

^aJC School of Public Health and Primary Care, the Chinese University of Hong Kong, Hong Kong SAR, ^bDepartment of Epidemiology and Biostatistics, Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, School of Public Health, Nanjing Medical University, Nanjing, P.R. China, ^cDepartment of Surgery, North District Hospital, ^dDepartment of Oncology, Princess Margaret Hospital, ^eDepartment of Surgery, Pamela Youde Nethersole Eastern Hospital, ^fDepartment of Pathology, North District Hospital, ^gDepartment of Medicine and Geriatrics, Princess Margaret Hospital, Kowloon, ^hDepartment of Pathology, Yan Chai Hospital, ⁱDepartment of Surgery, Yan Chai Hospital, Hong Kong SAR, ^jGenetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

* Correspondence: Lap Ah Tse, JC School of Public Health and Primary Care, the Chinese University of Hong Kong, Sha Tin, NT, Hong Kong SAR, China (e-mail: shelly@cuhk.edu.hk).

Copyright © 2016 the Author(s). Published by Wolters Kluwer Health, Inc. All rights reserved.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

Medicine (2016) 95:32(e4515)

Received: 10 January 2016 / Received in final form: 7 June 2016 / Accepted: 13 July 2016

<http://dx.doi.org/10.1097/MD.0000000000004515>

for developing efficient prevention strategies against BCa among Chinese women of Hong Kong.

The 1st step for BCa prevention is to accurately identify individuals at an increased BCa risk. Gail model is one of the most popular BCa risk assessment and prediction tools that has been widely used clinically to predict risk for individual women with Caucasian and African ethnic origin,^[5,6] although it was originally designed to identify women at an increased risk for entry onto chemoprevention trials and has limited discriminatory power (about 60% for the values of receiver-operator characteristic [ROC] curves).^[7,8] There are several other models developed to predict BCa risk in different populations by using similar indicators, such as the model from International Breast Cancer Intervention Study (IBIS) which focuses on familial BCa^[9] and the Mayo BBD-to-BCa model which focuses on women with history of benign breast diseases.^[10] Up till now, only 2 studies assessed the risk of BCa (Shanghai and Nanjing) in mainland China^[11,12] by including a limited number of SNPs and some risk factors, while the discriminatory power was also about 60% in these studies. The risk prediction model among Chinese women in Singapore also shows a similar accuracy.^[13] Furthermore, different fertility pattern, lifestyle, and environmental factors between mainland and Hong Kong also makes the direct application of prediction tools based on mainland Chinese to Hong Kong women a challenge.

Some established risk factors are known to vary by menopausal status^[14]; however, menopausal status is absent in most risk assessment models. Further, some potential risk factors identified in recent years (e.g., exposure to light at night [LAN], sleep disturbance, shift work with circadian disruption)^[15–18] were not addressed in the previous risk assessment models, which may lead to underestimated discriminatory power of these models. This study aims to construct a BCa risk assessment model specific to Hong Kong Chinese women, by including menopausal status and additional environmental risk factors that have never been considered in previous risk models.

2. Methods

2.1. Study subjects and specification of risk factors

The study protocol was approved by both the Clinical Research Ethics Committees of Joint Chinese University of Hong Kong – New Territories East Cluster (Joint CUHK-NTEC CREC) and the Kowloon West Cluster (KWC CREC). It was a hospital-based case–control study and the detailed process of subjects recruitment was described previously.^[19] In brief, eligible cases were consecutively identified from the Department of Surgery or Clinical Oncology of 3 hospitals in Hong Kong during the period November 2011 to March 2015. All included cases were Chinese women, aged between 20 and 84 with newly diagnosed primary BCa (ICD-10 code 50), and confirmed by histology. Specifically, we obtained the diagnostic results from either the tissue pathological or biopsy report for the cases receiving breast tissue dissection or biopsy, respectively. The response rate of cases was 90.8%, and main reasons for the nonresponses were no interest or poor medical condition. Each case was frequency matched in 5-year age groups by a control patient selected from the same hospital where the cases came from, with a response rate of 93.2%. Participants were excluded if they had a physician-diagnosed cancer history at any site.

After obtaining the written informed consent, face-to-face interviews were conducted by trained interviewers. A standard-

ized questionnaire was used to obtain the information on demographic data, menstrual and reproductive history (such as age at menarche, menopausal status, parity, breast feeding, age at 1st give birth, and hormone replacement therapy), history of benign breast diseases, family cancer history, occupational history, smoking habit, alcohol drinking, self-report exposure to LAN (1, dark; 2, few bright; 3, little bright; and 4, bright), and self-report sleep quality (1, good; 2, common; 3, poor; and 4, poor with sleep pill). The clinical information (including estrogen receptor [ER]) for the cases and controls was extracted from the hospital medical records.

2.2. Statistical analyses

We transformed relevant continuous variables to the categorical scale for the use of clinical practice.^[11] In the development of the risk model, we initially performed univariate logistic regression analyses separately by menopausal status to identify potential predictors. To include more variables in the initial selection, predictors with $P \leq 0.10$ were considered for further evaluation. Then, LASSO model was performed to select the potential predictors to be further used by shrinking the coefficients toward zero through setting a constraint on the sum of the absolute standardized coefficients. Shrinkage estimates with LASSO method in fitting model provided an important way for adjusting model's overfitting and preventing extreme predictions.^[20] The factors selected by LASSO selection were the predictors for the final model (LASSO-model). Third, multivariate logistic regression analyses were performed to generate the coefficients of all predictors from LASSO selection. We then assessed the effect of LASSO-model using a risk score analysis on the basis of a linear combination of the selected predictors weighted by their coefficients, using a formula as following: $S = \sum_{i=1}^n x_i * 10$, where S is the risk score, x the selected predictors, and n is the number of selected predictors. To facilitate result demonstration, the final risk score was multiplied by 10.

Ten-fold cross-validation was used to evaluate the model's internal validity by partitioning the original sample into a training set to train the model and a test set to evaluate it.^[21] The model performance was evaluated by ROC curves and the area under the curve (AUC) was used to classify the BCa cases and controls. Hosmer–Lemeshow fit test was used to assess the agreement between observed and model-predicted proportions of BCa events.^[22] The difference of AUCs was tested by a nonparametric approach developed by DeLong et al.^[23]

An optional model (OPT-model) which included the established risk factors of age, age at menarche, age at 1st give birth, number of BCa cases in 1st-degree relatives, and history of benign breast diseases was also constructed to compare the discrimination accuracy with LASSO-model. Their coefficients derived from multivariate logistic regression analysis were used as their category weightings. The discriminative power of LASSO-model was also evaluated among women in different age groups and different ER status of BCa. All statistical analyses were 2-sided and performed with Stata software (Version 11.2; StataCorp LP, TX).

3. Results

A total of 923 BCa cases and 918 age-matched controls were included in the final analysis. The characteristics of these subjects were summarized in Table 1. The mean age was 56.0 ± 11.8 years old in cases and 53.8 ± 11.8 years old in controls. BMI, age at

Table 1
Selected characteristics between breast cancer cases and controls of Hong Kong women.

Variables		Controls (n=918)	Cases (n=923)	P
Age	Mean ± SD	53.8 ± 11.8	56.0 ± 11.8	<0.001
	20–30 years	22 (2.4)	6 (0.7)	0.001
	30–40 years	87 (9.5)	70 (7.6)	
	40–50 years	246 (26.8)	232 (25.1)	
	50–60 years	319 (34.8)	317 (34.3)	
	60–70 years	159 (17.3)	169 (18.3)	
Body mass index	70~ years	85 (9.3)	129 (14.0)	
	<18.5	74 (8.1)	44 (4.8)	0.015
	18.5–24	570 (62.1)	567 (61.4)	
	24–27	160 (17.4)	175 (19.0)	
Age at menarche	27~	114 (12.4)	137 (14.8)	
	<12 years	127 (13.8)	99 (10.7)	0.005
	12-years	163 (17.8)	199 (21.6)	
	13-years	203 (22.1)	234 (25.4)	
	14-years	119 (13.0)	141 (15.3)	
	15-years	113 (12.3)	94 (10.2)	
Number of parity	16~ years	193 (21.0)	156 (16.9)	
	0	171 (18.6)	210 (22.8)	0.042
	1	175 (19.1)	168 (18.2)	
	2	300 (32.7)	316 (34.2)	
Age at 1st birth	3~	272 (29.6)	229 (24.8)	
	<20 years	56 (7.5)	43 (6.0)	0.029
	20–25 years	287 (38.4)	236 (33.1)	
Ever breast feeding	25~ years	148 (19.8)	180 (25.3)	
		343 (37.4)	303 (32.8)	0.041
Menopausal status	Pre-	351 (38.2)	341 (36.9)	0.568
	Post-	567 (61.8)	582 (63.1)	
Age at menopausal	<45 years	61 (10.8)	61 (10.5)	0.833
	45- years	95 (16.8)	82 (14.1)	
	49- years	216 (38.1)	229 (39.4)	
	51- years	72 (12.7)	73 (12.5)	
	53- years	62 (10.9)	73 (12.5)	
	55~ years	61 (10.8)	64 (11.0)	
Oral contraceptive		333 (36.3)	301 (32.6)	0.098
Hormone replace treatment		35 (3.8)	19 (2.6)	0.026
Case number of BCFDR	0	889 (96.8)	953 (92.4)	<0.001
	1	27 (2.9)	61 (6.6)	
	2	2 (0.2)	9 (1.0)	
Benign breast diseases		271 (29.5)	296 (32.1)	0.236
Ever smoking		69 (7.5)	63 (6.8)	0.566
Frequently alcohol drinking		41 (4.5)	47 (5.1)	0.529
Light at night	Dark	491 (53.5)	424 (45.9)	0.005
	Few bright	384 (41.8)	451 (48.9)	
	Little bright	43 (4.7)	48 (5.2)	
Sleep quality	Good	286 (31.2)	332 (36.0)	0.077
	Common	377 (41.1)	374 (40.5)	
	Poor	206 (22.4)	180 (19.5)	
	Very poor	49 (5.3)	37 (4.0)	
Estrogen receptor status	Negative	183 (23.3)	—	
	Positive	601 (76.7)	—	

BCFDR=breast cancer in first degree-relatives, SD=standard deviation.

menarche, number of parity, age at 1st give birth, ever breast feeding, hormone replace treatment, BCa history among 1st-degree relatives, and exposure to LAN were significantly different between cases and controls. After stratified by menopausal status, the alcohol drinking rate was significantly higher in premenopausal cases of BCa. The distribution of number of parity, breast feeding, and hormone replace treatment were significantly different between cases and controls among postmenopausal women (Supplemental Table 1, <http://links.lww.com/MD/B199>). There were 784 BCa patients with data on ER status, and among

them, 601 (76.7%) cases were ER positive (H-score > 50, or Allred score ≥ 3).

The potential predictors selected by univariate logistic regression are presented in Table 2 and supplemental Table 2, <http://links.lww.com/MD/B199>. These potential predictors were refined by LASSO selection (Table 2), which were generated by multivariate logistic regression model that were comparable to the OPT-model (supplemental Table 2, <http://links.lww.com/MD/B199>). Eventually, 6 risk factors were included in the final LASSO model for premenopausal women and 12 factors were

Table 2

Predictors for breast cancer by LASSO selection and their coefficient calculated by multivariate logistic regression model.

Variables	Coefficient	
	Premenopausal	Postmenopausal
Age	0.367	0.346
Body mass index		0.935
Age at menarche		-0.075
Age at 1st give birth		0.141
Number of parity	0.137	-0.184
Breast feeding		-0.110
Oral contraceptive		-0.090
hormone replace treatment		-0.710
Case number of BCFDR	0.855	0.844
Benign breast diseases		0.296
Alcohol drinking	0.631	
LAN	0.264	0.238
Sleep quality	-0.256	-0.122

Age (20, 30, 40, 50, 60, 70, and >70 years old); body mass index (<18.5, 18.5–24, 24–27, and ≥27); age at menarche (<12, 12, 13, 14, 15, and 16~ years old); age at 1st give birth (<20, 20–25, and 25~ years old); number of parity (0, 1, 2, and >2); breast feeding duration (no, <1, 1–3 and, >3 years); LAN (1, dark; 2, few light; and 3, little bright); sleep quality (1, good; 2, common; 3, poor; and 4, poor with sleep pill). BCFDR=breast cancer in first degree-relatives, LAN=light at night, LASSO=least absolute shrinkage and selection operator, SD=standard deviation.

included for postmenopausal women. Age, number of parity, case number of BCa in first-degree relatives, LAN, and sleep quality were the common predictors for both pre- and postmenopausal women. The coefficient for most variables was similar in both groups of women, except for number of parity (0.137 and -0.184 for pre- and postmenopausal women, respectively). In addition to these factors, alcohol drinking was included for premenopausal women; BMI, age at menarche, age at 1st give birth, ever breast feeding, ever using of oral contraceptive, hormone replacement treatment, and history of benign breast diseases were included for postmenopausal women.

Figure 1 demonstrates the results of risk score analysis from the LASSO-model incorporating with a linear combination of the risk predictors weighted by their coefficients. The means of the risk scores was 7.43 ± 4.63 for controls and 9.74 ± 5.05 for BCa cases among premenopausal women, and 10.44 ± 5.32 for controls and 13.60 ± 5.97 for cases among postmenopausal women. The discrimination accuracy of this score system was determined by the ROC curves (Fig. 2), given the AUC of 0.640 (95% CI, 0.598–0.681) for premenopausal women and 0.655 (95% CI, 0.623–0.686) for postmenopausal women, respectively. Hosmer–Lemeshow fit test showed good agreement between observed and model-predicted proportions of BCa in both pre- and postmenopausal women ($P=0.302$ and 0.848 , respectively).

A 10-fold cross-validation method was applied to split all the samples randomly into 10 partitions, and taking 9-fold as training set, and another 1-fold as the validation set, 1000 times were repeated to evaluate the OPT-model's ability. The risk score calculated by the coefficients of the predictors was evaluated with each random datasets, and the results were similar with the observed data (average AUC: 0.621 and 0.632 for pre- and postmenopausal women, respectively). These simulation results showed a robust internal consistence between the estimated effects of predictors based on the original results and the bootstrapped results.

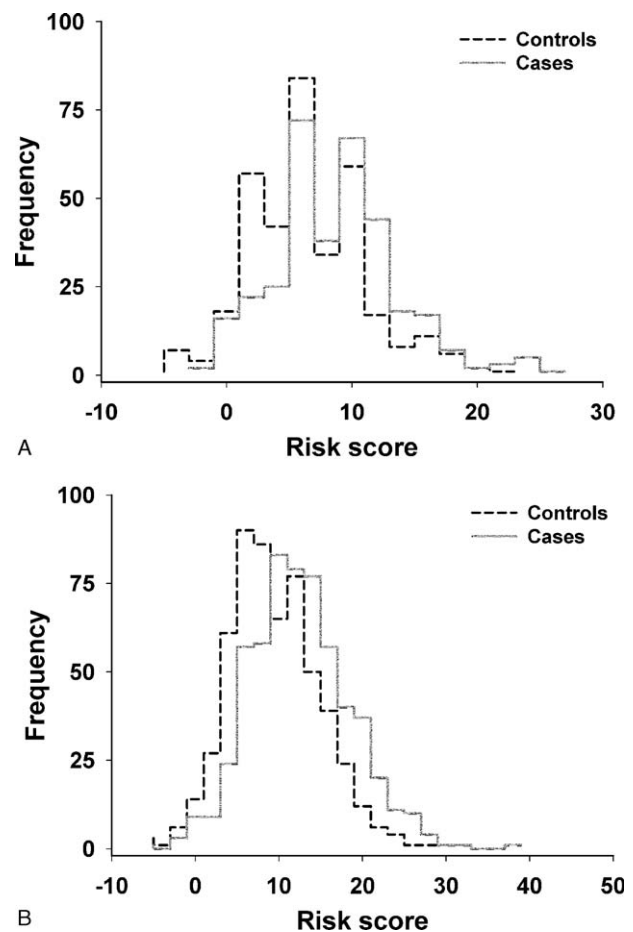


Figure 1. Distribution of risk scores calculated with a linear combination of the least absolute shrinkage and selection operator (LASSO) selected predictors weighted by multivariate logistic regression coefficient. (A) Premenopausal women; (B) postmenopausal women.

Compared to the OPT-model, the discrimination accuracy of LASSO-model improved significantly (Fig. 2). The AUCs were increased from 0.586 to 0.640 among premenopausal women ($P=0.011$) and from 0.621 to 0.655 among postmenopausal women ($P=0.006$), respectively. The comparison of discrimination accuracy among women with different ages is presented in Table 3. This score system showed better accuracy among postmenopausal women aged between 50 and 70 years old, and among premenopausal women aged between 40 and 50 years. For ER positive BCa, the accuracy was higher in postmenopausal women with AUC of 0.663 (95% CI, 0.628–0.696) (Supplemental Table 3, <http://links.lww.com/MD/B199>).

4. Discussion

There was no risk assessment tool specifically for identifying Hong Kong women at increased risk of developing BCa prior to this study. We developed a risk assessment model for invasive BCa of Hong Kong women. By adding modifiable risk factors (e.g., exposure to LAN and sleep quality) that had never been addressed in previous risk models, the discrimination accuracy significantly improved from 0.586 to 0.640 among premenopausal women and from 0.621 to 0.655 among postmenopausal women, and a better model was demonstrated for the

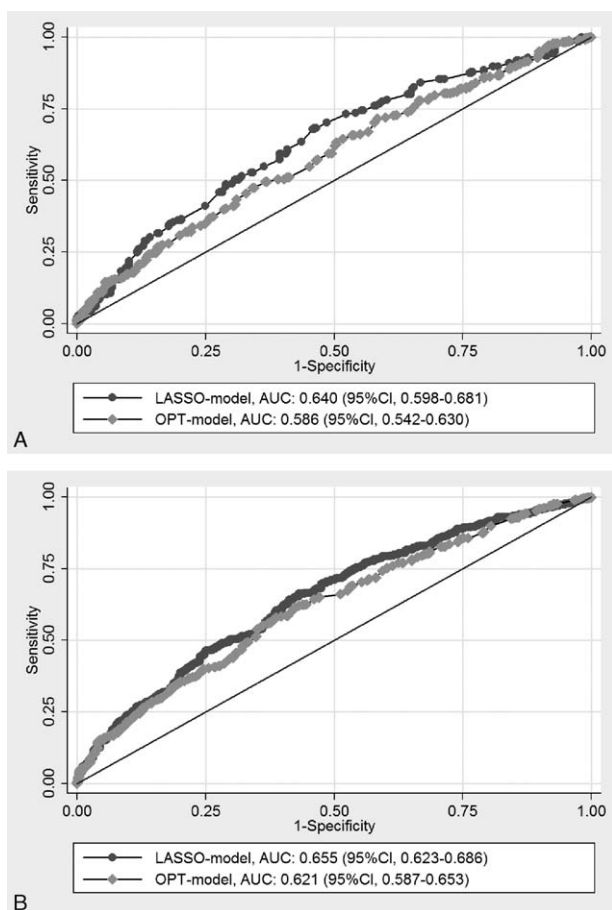


Figure 2. Discrimination performance of LASSO-model for breast cancer and comparison with OPT-model by receiver-operator characteristic analyses among pre- and postmenopausal Hong Kong women (AUC). (A) Premenopausal women; (B) postmenopausal women. AUC=area under the curve, LASSO=least absolute shrinkage and selection operator, OPT-model=optional model, ROC=receiver-operator characteristic.

postmenopausal women aged between 50 and 70 years and for the ER positive BCa. Overall, the performance of this model is slightly superior to other risk models in which the discriminatory power is generally around 60%.

In our study, only 5 of 12 predictors in the risk model of postmenopausal women were enrolled in the risk model of premenopausal women, and the effect of parity was opposite between pre- and postmenopausal women (coefficient were 0.137

and -0.184 , respectively). These results suggest that the influence of menopause is rather comprehensive than a single event for the initiation of BCa. Hormones produced by the ovaries play central roles in breast tissue development, maintenance, and tumorigenesis.^[14,24] However, menopause changes the existing circumstance which might affect the entire risk profile of BCa. For example, obesity is associated with BCa incidence, but it showed different risk patterns according to the women’s menopausal status.^[25,26] The study of KoBRCAT observed the same phenomenon as our study.^[27] The major risk factors and their effects in KoBRCAT were different after dividing the participants into 2 subgroups by age of 50 years old. Our study is able to attribute this variety in terms of risk factors to the etiology of BCa incidences before and after menopause, and thus provides scientific basis for decision makers to resettle the proper role of menopause in BCa risk assessment and prediction.

Gail model was used to predict the individual risk for developing BCa in western population.^[28] However, although recalibrations have been made, it does not perform well in Asian women. The recalibrated Gail model performed good calibration in the study of Singapore Breast Screening Programme (model name, Gail-SBSP) and Seoul Breast Cancer Study (model name, KoBRCAT), but the discriminative power was limited with the AUC of about 0.60.^[13,27] After involving modifiable risk factors of BMI, oral contraceptive usage, and exercise, the discriminative power of KoBRCAT improved significantly to 0.63 for women <50 years old and 0.65 for women ≥ 50 years old, respectively. In the current study, we enrolled more predictors in the risk model including BMI, breast feeding history, LAN, sleep quality and alcohol drinking, and the discriminative accuracy was significantly improved.

It is the 1st time that LAN exposure was enrolled into the BCa risk assessment model. There is a theoretical link between LAN and BCa occurrence.^[16] Long-term exposure to LAN can suppress the secretion of hormone melatonin and cause the circadian disruption.^[29] Melatonin appears to protect against cancer development,^[30] and decreased secretion of melatonin may induce continuous production of estrogen and alter the function of ER.^[31] There is no epidemiological report which focused on LAN and BCa risk. However, as an extreme scenario of exposure to LAN, night shift work showed positive association with BCa risk.^[32] Except for the working circumstance with bright light, mechanism study showed a significantly suppressed of nocturnal melatonin secretion by very low illuminance of blue-appearing light.^[33] In this study, the brightness might be very low in bedroom during the sleep time and we did not assess the light levels objectively. But this result warrants further studies on LAN exposure and the risk of BCa.

Table 3

Prediction accuracy of risk scores among different age groups in Hong Kong women.

Age group	Premenopausal women				Postmenopausal women			
	Subjects	AUC	95% CI	P*	Subjects	AUC	95% CI	P*
<30 years	27	0.480	0.207–0.753	0.305	—	—	—	0.559
30–40 years	150	0.553	0.459–0.647		—	—	—	
40–50 years	395	0.637	0.583–0.691		83	0.611	0.487–0.736	
50–60 years	116	0.573	0.462–0.684		520	0.641	0.594–0.689	
60–70 years	—	—	—		325	0.683	0.625–0.740	
≥ 70 years	—	—	—	213	0.627	0.551–0.703		

AUC=area under the curve, CI=confidence interval.

* P value for the comparison among the AUCs of different age groups.

Three studies reported the relationship between sleep quality and risk of BCa,^[34–36] but no association was found; however, we observed a negative relation between poor sleep quality and BCa risk, and the underlying reason remains unclear. We tried to exclude those with chronic comorbidities (may potentially influence sleep quality) from both cases and controls, and results remain unchanged. Further studies with prospective study design are needed to investigate the role of sleep quality and BCa occurrence.

BMI, which was used as the indicator for obesity, was included into our postmenopausal women model as well as in the KoBRCA model for women aged older than 50 years. BMI was not selected into the model for premenopausal women in our study or for young women in KoBRCA, although it was reported to have a protective effect among premenopausal women.^[25,26] Considering potential differences in the manifestation of weight gain between Asian and western women, waist circumference may be a better indicator that should be addressed in future studies.

Alcohol drinking has long been shown to be associated with an increased risk of BCa.^[37] Every 10g of alcohol consumption per day might increase the risk up to 12%.^[38,39] In this study, alcohol drinking was involved into the risk assessment model for premenopausal women with high weight in the score system. However, it was excluded from the model for postmenopausal women. The possible reason is that the drinking rate is higher in young women than old women in Hong Kong, according to a recent study of BCa in Hong Kong.^[40]

Our BCa risk assessment model showed better performance for ER positive BCa and women aged between 50–70 years. It is easy to understand the better accuracy for ER positive BCa, because most selected factors are directly or indirectly associated with previous estrogen levels. The lower discriminative power for women older than age 70 years might be the results of a weaker association of risk factors with BCa risk among older women^[41] and fewer participants.

There are several strengths of this study. First, we constructed the risk assessment model by menopausal status which might more clearly reflect the risk changes before and after stopping menstruation. Second, several modifiable risk factors were involved in the risk model, which might be used for the primary prevention of BCa. Third, our study is the 1st to provide a potential value of using LAN exposure to predict the risk of BCa. Nevertheless, several limitations should be mentioned. First, this study used a case–control study design and recall bias might be a concern. We compared information on a special group of 117 patients (not including in this study) who were handled as BCa initially but finally confirmed to be noncases, and found that the prevalence of various risk factors were slightly lower than the true cases, which suggested the information bias may not be a major methodological issue. And we also conducted test–retest reliability among 25% cases and controls after 6 weeks of the initial interview. The kappa value was 0.62 which indicates a relatively high reliability and low misclassification of the potential variables. Therefore, the issues of residual confounding effect and reliability of the potential risk predictors in our study should be low. Second, the study design did not allow us to capture the changes of some modifiable risk factors such as BMI and sleep quality through menopause. These changes need to be investigated in further study with cohort study design. Third, we recruited controls from hospitals rather than from general population, and some exposures may differ between hospital patients and the general population. To reduce the potential bias caused by using hospital-based

controls, we excluded controls with breast-related diseases and recruited controls with a broad spectrum of diagnosis. Fourth, germline mutation of BRCA1/2 status was unavailable to us; however, this should not influence the risk assessment significantly as the prevalence of germline BRCA1/2 mutations is very low (1%) in Hong Kong.^[42] Also genetic variants and mammographic density were unavailable for this study. They are promising factors for BCa prediction, although limited contribution was observed in previous studies.^[43] Fifth, we did not collect potential risk factors' changes as the subject progress through menopause, because the potential misclassification from the recall of previous environmental exposures in different time points may be substantial, especially for the nonhabitual exposures. Finally, all BCa patients came from 3 public hospitals that may not represent all cases in the entire population; however, our case samples are comparable to those obtained from the Hong Kong Cancer Registry in which a wide coverage of approximate 95% of the Hong Kong general population is indicated regarding age and histological subtypes.^[44]

In conclusion, considering the rapid increase of BCa incidence in recent decades, a better risk assessment model specific to Hong Kong women is urgently needed for assessing and predicting BCa risk. We developed a novel risk assessment model by including menopausal status and emerging environmental risk factors that have never been explored in previous models, and demonstrated better discriminative accuracy than previous risk models in other populations; hence, results from our study have added new scientific evidence to the current literature in terms of risk prediction on BCa. We expect that newly developed model shall be used for the screening of high risk population of BCa and contribute to primary and secondary prevention of BCa in Hong Kong.

Acknowledgements

The authors thank Research Grants Council of Hong Kong (Grant number 474811) for the support. The authors also thank Miss Yin-shan Magdalene Leung, Hung-kuen Ivy Hsu, Kit-ping Apple Kwok, and Angela Chan for their assistance in patients' recruitment and data collection.

References

- [1] Hong Kong Cancer Registry. Female Breast Cancer in 2012. HongKong: Hong Kong Cancer Registry, Hospital Authority; November, 2014.
- [2] DeSantis C, Ma J, Bryan L, et al. Breast cancer statistics, 2013. *CA Cancer J Clin* 2014;64:52–62.
- [3] Hong Kong Cancer Registry. <http://www3.ha.org.hk/cancereg/statistics.html>. (Accessed August 6, 2015).
- [4] Sung H, Rosenberg PS, Chen WQ, et al. Female breast cancer incidence among Asian and Western populations: more similar than expected. *J Natl Cancer Inst* 2015;107: pii: djv107.
- [5] Gail MH, Costantino JP, Pee D, et al. Projecting individualized absolute invasive breast cancer risk in African American women. *J Natl Cancer Inst* 2007;99:1782–92.
- [6] Costantino JP, Gail MH, Pee D, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst* 1999;91:1541–8.
- [7] Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;81:1879–86.
- [8] Rockhill B, Spiegelman D, Byrne C, et al. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst* 2001;93:358–66.
- [9] Cuzick J, Forbes J, Edwards R, et al. First results from the International Breast Cancer Intervention Study (IBIS-I): a randomised prevention trial. *Lancet* 2002;360:817–24.

- [10] Pankratz VS, Degnim AC, Frank RD, et al. Model for individualized prediction of breast cancer risk after a benign breast biopsy. *J Clin Oncol* 2015;33:923–9.
- [11] Dai J, Hu Z, Jiang Y, et al. Breast cancer risk assessment with five independent genetic variants and two risk factors in Chinese women. *Breast Cancer Res* 2012;14:R17.
- [12] Wang Y, Gao Y, Battsend M, et al. Development of a risk assessment tool for projecting individualized probabilities of developing breast cancer for Chinese women. *Tumour Biol* 2014;35:10861–9.
- [13] Gao F, Machin D, Chow KY, et al. Assessing risk of breast cancer in an ethnically South-East Asia population (results of a multiple ethnic groups study). *BMC Cancer* 2012;12:529.
- [14] Eden JA. Breast cancer, stem cells and sex hormones. Part 3: the impact of the menopause and hormone replacement. *Maturitas* 2011;68:129–36.
- [15] Trinh T, Christensen SE, Brand JS, et al. Background risk of breast cancer influences the association between alcohol consumption and mammographic density. *Br J Cancer* 2015;113:159–65.
- [16] Stevens RG, Brainard GC, Blask DE, et al. Breast cancer and circadian disruption from electric lighting in the modern world. *CA Cancer J Clin* 2014;64:207–18.
- [17] Straif K, Baan R, Grosse Y, et al. Carcinogenicity of shift-work, painting, and fire-fighting. *Lancet Oncol* 2007;8:1065–6.
- [18] Gao Y, Huang YB, Liu XO, et al. Tea consumption, alcohol drinking and physical activity associations with breast cancer risk among Chinese females: a systematic review and meta-analysis. *Asian Pac J Cancer Prev* 2013;14:7543–50.
- [19] Tse LA, Li M, Chan WC, et al. Familial risks and estrogen receptor-positive breast cancer in Hong Kong Chinese women. *PLoS One* 2015;10:e0120741.
- [20] Steyerberg E. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. 1 ed. New York: Springer-Verlag New York; 2009.
- [21] Efron B. *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia, Pennsylvania, USA: Society for Industrial and Applied Mathematics; 1982.
- [22] Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York: Wiley; 1989.
- [23] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–45.
- [24] Eden JA. Human breast cancer stem cells and sex hormones – a narrative review. *Menopause* 2010;17:801–10.
- [25] Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. *Lancet Oncol* 2001;2:133–40.
- [26] Veronesi U, Boyle P, Goldhirsch A, et al. Breast cancer. *Lancet* 2005;365:1727–41.
- [27] Park B, Ma SH, Shin A, et al. Korean risk assessment model for breast cancer risk prediction. *PLoS One* 2013;8:e76736.
- [28] Mahoney MC, Bevers T, Linos E, et al. Opportunities and strategies for breast cancer prevention through risk reduction. *CA Cancer J Clin* 2008;58:347–71.
- [29] Costa G, Haus E, Stevens R. Shift work and cancer – considerations on rationale, mechanisms, and epidemiology. *Scand J Work Environ Health* 2010;36:163–79.
- [30] Blask DE. Melatonin, sleep disturbance and cancer risk. *Sleep Med Rev* 2009;13:257–64.
- [31] Stevens RG, Blask DE, Brainard GC, et al. Meeting report: the role of environmental lighting and circadian disruption in cancer and other diseases. *Environ Health Perspect* 2007;115:1357–62.
- [32] Wang F, Yeung KL, Chan WC, et al. A meta-analysis on dose-response relationship between night shift work and the risk of breast cancer. *Ann Oncol* 2013;24:2724–32.
- [33] Glickman G, Levin R, Brainard GC. Ocular input for human melatonin regulation: relevance to breast cancer. *Neuro Endocrinol Lett* 2002;23 (Suppl 2):17–22.
- [34] Vogtmann E, Levitan EB, Hale L, et al. Association between sleep and breast cancer incidence among postmenopausal women in the Women's Health Initiative. *Sleep* 2013;36:1437–44.
- [35] Girschik J, Heyworth J, Fritschi L. Self-reported sleep duration, sleep quality, and breast cancer risk in a population-based case-control study. *Am J Epidemiol* 2013;177:316–27.
- [36] Verkasalo PK, Lillberg K, Stevens RG, et al. Sleep duration and breast cancer: a prospective cohort study. *Cancer Res* 2005;65:9595–600.
- [37] Baan R, Straif K, Grosse Y, et al. Carcinogenicity of alcoholic beverages. *Lancet Oncol* 2007;8:292–3.
- [38] Allen NE, Beral V, Casabonne D, et al. Moderate alcohol intake and cancer incidence in women. *J Natl Cancer Inst* 2009;101:296–305.
- [39] Chen WY, Rosner B, Hankinson SE, et al. Moderate alcohol consumption during adult life, drinking patterns, and breast cancer risk. *JAMA* 2011;306:1884–90.
- [40] Yeo W, Lee HM, Chan A, et al. Risk factors and natural history of breast cancer in younger Chinese women. *World J Clin Oncol* 2014;5:1097–106.
- [41] Vacek PM, Skelly JM, Geller BM. Breast cancer risk assessment in women aged 70 and older. *Breast Cancer Res Treat* 2011;130:291–9.
- [42] Cancer Expert Working Group on Cancer Prevention and Screening. *Recommendations on Breast Cancer Screening*. Hong Kong: Cancer Expert Working Group on Cancer Prevention and Screening, Hong Kong SAR Government; November, 2012.
- [43] Vachon CM, Pankratz VS, Scott CG, et al. The contributions of breast density and common genetic variation to breast cancer risk. *J Natl Cancer Inst* 2015;107: pii: dju397.
- [44] Tse LA, Li M, Chan WC, et al. Familial risks and estrogen receptor-positive breast cancer in Hong Kong Chinese women. *PLoS One* 2015;10:e0120741.