Original Article

# Initiators and promoters for the occurrence of screen-detected breast cancer and the progression to clinically-detected interval breast cancer

Amy Ming-Fang Yen [a, 1], Wendy Yi-Ying Wu [b, c, 1], Laszlo Tabar [d], Stephen W. Duffy [e], Robert A. Smith [f], Hsiu-Hsi Chen [b, *]

[a] School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan
[b] Graduate Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan
[c] Department of Radiation Sciences, Umeå University, Umeå, Sweden
[d] Department of Mammography, Falun Central Hospital, Falun, Sweden
[e] Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK
[f] Cancer Control Science Department, American Cancer Society, Atlanta, GA, USA

ABSTRACT

Background: The risk factors responsible for breast cancer have been well documented, but the roles of risk factors as initiators, causing the occurrence of screen-detected breast cancer, or promoters, responsible for the progression of the screen-detected to the clinically-detected breast cancer, have been scarcely evaluated.
Methods: We used data from women in a cohort in Kopparberg (Dalarna), Sweden between 1977 and 2010. Conventional risk factors, breast density, and tumor-specific biomarkers are superimposed to the temporal course of the natural history of the disease.
Results: The results show that older age at first full-term pregnancy, dense breast, and a family history of breast cancer increased the risk of entering the preclinical screen-detectable phase of breast cancer by 23%, 41%, and 89%, respectively. Overweight/obesity (body mass index $\geq$25 kg/m$^2$) was a significant initiator (adjusted relative risk [aRR] 1.15; 95% confidence interval [CI], 0.99–1.33), but was inversely associated with the role of promoter (aRR 0.65; 95% CI, 0.51–0.82). Dense breast (aRR 1.46; 95% CI, 1.12–1.91), triple-negative (aRR 2.07; 95% CI, 1.37–3.15), and Ki-67 positivity (aRR 1.66; 95% CI, 1.19–2.30) were statistically significant promoters. When the molecular biomarkers were considered collectively as one classification, the basal-like subtype was the most influential subtype on promoters (aRR 4.24; 95% CI, 2.56–7.02) compared with the Luminal A subtype.
Discussion: We ascertained state-dependent covariates of initiators and promoters to classify the risk of the two-step progression of breast cancer. The results of the current study are useful for individually-tailored screening and personalized clinical surveillance of patients with breast cancer that was detected at an early stage.
© 2016 The Authors. Publishing services by Elsevier B.V. on behalf of The Japan Epidemiological Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

## 1. Introduction

Hormonal risk factors that are responsible for breast cancer have been well documented since 1980.[1,2] The majority of studies place emphasis on whether or not breast cancer occurs. Mathematical models that predict the risk of breast cancer, such as the Gail model, have been proposed for such a purpose.[3–7] In the era of preventive medicine, a simple relationship of a particular risk factor to the occurrence of breast cancer is not sufficient. Considering the

* Corresponding author. Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Room 533, No. 17, Hsuchow Road, Taipei, Taiwan.
E-mail address: chenlin@ntu.edu.tw (H.-H. Chen).
Peer review under responsibility of the Japan Epidemiological Association.
[1] These authors made an equal contribution.
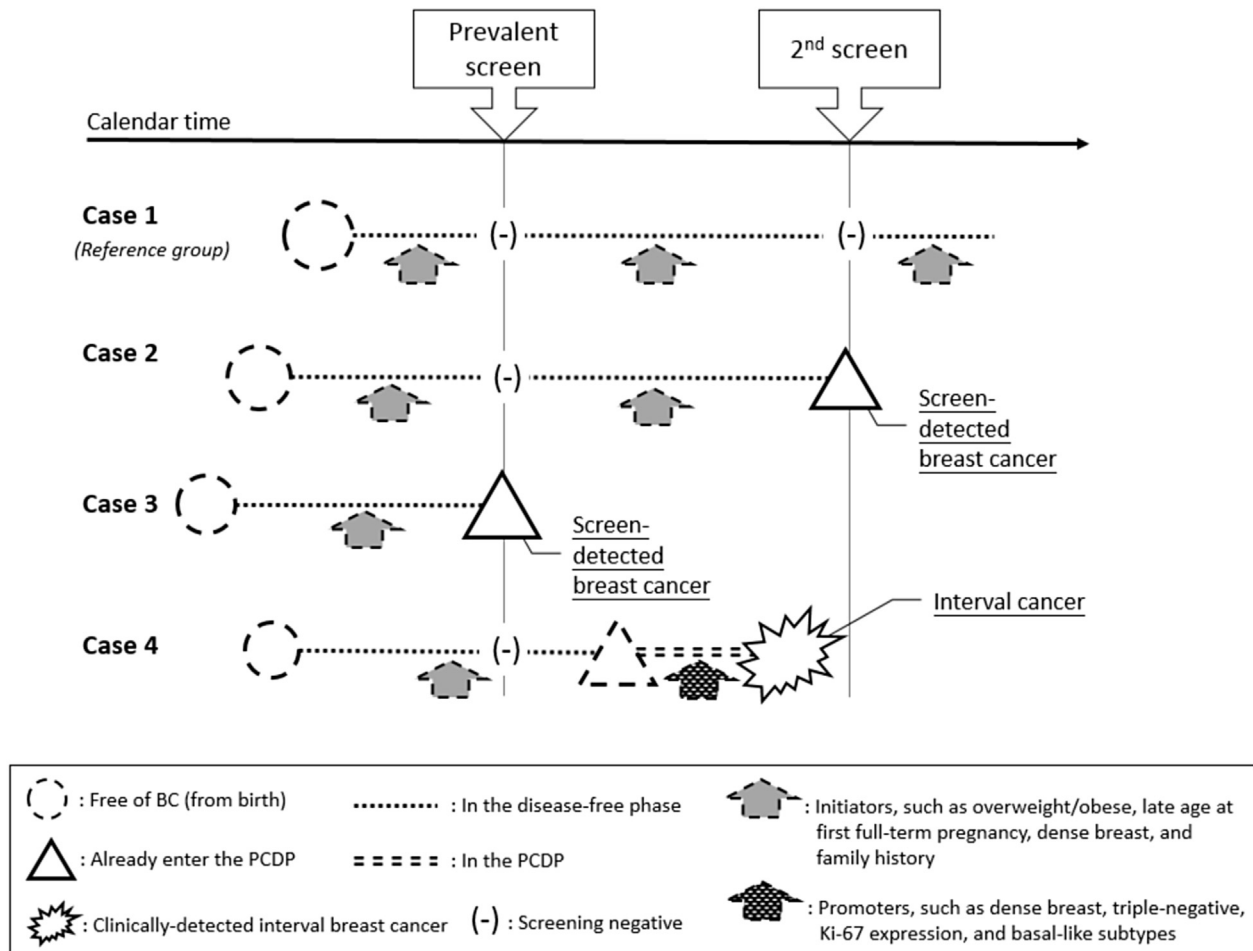
**Fig. 1.** Breast tumors that remained in the PCDP and the CP, as represented by screen-detected and clinically-detected interval breast cancers in relation to initiators and promoters. * The dashed line represents the unobserved status, while the solid line and underscored text represent the observed status. BC, breast cancer; CP, clinical phase; PCDP, pre-clinical detectable phase.

dynamic process of the development of breast cancer as shown in Fig. 1, women experience a pre-symptomatic phase before they progress to the symptomatic phase. The majority of women are initially free of breast cancer (case 1). In cases of breast cancer, the tumor develops as a very small and pre-symptomatic lesion, which is undetectable. As time passes, the tumors clone and grow to a detectable size, and although women may still be pre-symptomatic during this phase, the tumor can be detected using available screening tools, such as mammography; even at this stage, women may still not exhibit any symptoms and signs (the so-called pre-clinical detectable phase [PCDP]). The PCDP is the window of early detection. If the lesion is detected through screening during the period when women are in the PCDP, they are defined as "screen-detected cases" (case 2 and case 3). If women enter the PCDP after screening but progress to a clinical phase after they feel lumps or experience symptoms and signs (the clinical phase [CP]) and then seek medical care, these are defined as "interval cancers" (case 4). The risk factors of interest may be related to the rate of entering the PCDP (initiators) and also may be responsible for the subsequent progression from the PCDP to the CP (promoters). The elucidation of the function of each risk factor with respect to its role as an initiator or a promoter is of great importance to individually-tailored screening and personalized clinical surveillance.

However, the classification of each risk factor as an initiator or a promoter is a great challenge unless the data from a population-based breast cancer screening are considered. These data provide an opportunity to assess the relative contribution between initiators and promoters through a comparison of the distribution of each risk factor between screen-detected breast cancers (pre-symptomatic cases that remained in the PCDP) and clinically-detected breast cancers. An example of clinically-detected breast cancer is interval cancer, which represents progression of cancer from the PCDP that was already found at a previous screen and was missed or that progressed to the CP after the screen but before subsequent screening (i.e., symptomatic cases). In addition to the conventional hormonal risk factors, information on the status of estrogen receptor (ER), progesterone receptor (PR), HER-2/neu, and Ki-67, as well as the presence of a basal-like phenotype, are widely used to predict the prognosis of breast cancer; these additional factors are also very informative with respect to the rate of progression from the pre-symptomatic phase to the symptomatic phase.[8−10] Because only breast tumor cases have such information, it is postulated that if the distributions of these factors are different between screen-detected and clinically-detected interval cancers, these tumor-specific markers might play a crucial role as promoters.

In the current study, we aimed to use longitudinal follow-up data from Kopparberg (Dalarna) county in the Swedish two-county trial of mammography screening. We applied a four-state continuous-time Markov regression model to estimate the

incidence rate of preclinical breast cancer based on screen-detected breast cancers and the transition rate from the PCDP to the CP based on interval breast cancers, taking into account the hazard rate of dying from causes other than breast cancer. We also estimated the relative risks (RRs) of the associated risk factors on the two rates governing the disease progression described above.

## 2. Methods

### 2.1. Study subjects

Subjects were enrolled from one county of the Swedish two-county trial in Kopparberg (Dalarna), with a follow-up period from 1977 until 2010. The details of the trial have been described in full elsewhere.[11] Briefly, women aged 40–74 years in the screening arm were invited to participate in screenings every 24–33 months. At the close of the trial (1985), the service screening with two-view mammography screening was recommended by the Swedish Board of Health and Welfare for women aged 40–69 years. Data on the individual screening history of patients, including negative findings, breast cancers detected during screens, and those cases that were diagnosed between the scheduled screens due to the occurrence of symptoms and signs, were collected and followed until 1992. Data on breast cancer cases from 1993 to 2010 from the cancer registry (which is the most common scenario in the real world) were used. Data on date and causes of death for all breast cancer patients were obtained from the death registry. The number of disease-free women after 1992 was approximated based on the population statistics in Sweden (http://www.scb.se). The average inter-screening interval after 1992 was 24 months. To estimate the hazard rate of competing causes of death other than breast cancer, frequencies of death of the underlying population by age and calendar year were retrieved from the population statistics in Sweden.

### 2.2. Study design

In all, 50,666 women aged 40–74 years at the beginning of study were identified. As conventional risk factors were not collected after 1992, and because tumor-specific factors were also collected in different calendar years after 1992, we divided the follow-up time into two eras: 1977–1992 and 1993–2010. The record of the participants in each screen was kept until 1992, as were data on the date of mammography, the screening results, and whether a clinical diagnosis was made after negative findings. In the current analysis, we did not include those who had never attended a screen but were diagnosed with clinically-detected breast cancer due to the presence of symptoms and signs (those who refused mammography). They were excluded because information on conventional risk factors was not available for the women who refused to attend the screen. Within the analytic cohort, we further distinguished each woman as breast cancer-free, as a case of screen-detected breast cancer (in the PCDP), or as a case of interval cancer (in the CP). During this period, 1321 breast cancers were identified, including 789 screen-detected and 532 interval cancers. Between 1993 and 2010, 1614 breast cancers (1135 screen-detected and 469 interval cancers) were diagnosed in women aged 40–69 years and were reported in the cancer registry. As mentioned earlier, the number of disease-free women at each round of screening after 1992 was approximated from the population statistics. A flowchart of the screening and the corresponding data is shown in Fig. 2. In the current analysis of four-state Markov model, the layout of data used for analysis is displayed in eTable 1, including vital status on death from causes other than breast cancer and for women free of breast cancer by 1-year age retrieved from the Swedish population statistics. Data on women-years of follow-up by 1-year age for women

in the PCDP and those free of breast cancer are also given in eTable 1.

### 2.3. Data collection

#### 2.3.1. Conventional risk factors

In the trial period, the screening staff took anthropometric measurements for each woman at the first screening. We used a body mass index (BMI) of 25 kg/m$^2$ to categorize women as overweight/obese (BMI $\geq$25 kg/m$^2$) or not (BMI <25 kg/m$^2$). A questionnaire was also used to obtain information on the age at menarche, menopausal status, age at first full-term pregnancy (AP), and family history of breast cancer.

#### 2.3.2. Breast density

All mammographs were performed by well-trained technicians from Falun Central Hospital following a standard procedure.[12] Women at the inception of the trial were classified with either baseline dense breast (Tabar patterns IV and V) or non-dense breast (Tabar patterns I-III).[13] Note that Tabar patterns IV and V correspond to Wolfe patterns P2 and DY, excluding QDY.[14]

#### 2.3.3. Tumor-specific biomarkers

All tumor-specific biomarkers were examined in the Department of Pathology, Falun Central Hospital starting in 1996. Biomarkers included ER, PR, HER-2, Ki-67, and basal/myoepithelial markers, but different biomarkers were available in different years depending on the hospital policy (see Fig. 2). Tumors that expressed at least one of the basal/myoepithelial markers, including cytokeratin (CK)5/6, CK14 and Epidermal Growth Factor Receptor, were classified as tumors with a basal-like phenotype. Immunohistochemistry-based molecular analysis allowed for further categorization into five molecular subtypes: Luminal A (ER+, PR+/−, HER-2−), Luminal B (ER+, PR+/−, HER-2+), HER-2-like (ER−, PR−, HER-2+), basal-like subtype (positive for at least one basal/myoepithelial marker) and triple-negative (ER−, PR−, HER-2−).[15] Note that because we treated the basal-like subtype as the dominate subtype, this subtype also included tumors with a basal-like phenotype. The other four molecular subtypes collectively included tumors without a basal-like phenotype. The data used for academic purposes have been approved by the Regional Ethics Committee in Uppsala, Örebro, Sweden (Dnr 2008/081).

#### 2.3.4. Statistical analysis

We applied a four-state continuous-time Markov model to delineate the disease progression of breast cancer from the breast cancer-free phase (state 0), to the PCDP (state 1), and finally, to the CP (state 2), making allowance for deaths due to causes other than breast cancer (state 3) from women with states 0 and 1 (see Fig. 3). Both states 2 and 3 are viewed as absorbing states, as the CP is the destination of the natural history of breast cancer and women free of breast cancer and in the PCDP may die from causes other than breast cancer. The methods of the multi-state Markov model have been well developed and have been previously applied to breast cancer screens.[16–20] The two instantaneous transition rates that govern the disease progression of breast cancer are denoted by $\lambda_1$ and $\lambda_2$ (Fig. 3). The mean sojourn time (MST) is equal to $1/\lambda_2$ under the exponential assumption. The hazard rates of other causes of death from states 0 and 1 are denoted by $\mu_0$ and $\mu_1$, respectively. This four-state model was further used to model the effect of covariates on different transitions. The log-linear hazards regression form was used to relate covariates to the incidence of breast cancer (transition rate from the breast cancer-free phase to the PCDP) and the transition from the PCDP to the CP.[10] The details of the statistical model are given in eAppendix 1. Those factors that

**Timeline**    **Flowchart of Screen**    **Covariates & Availability**
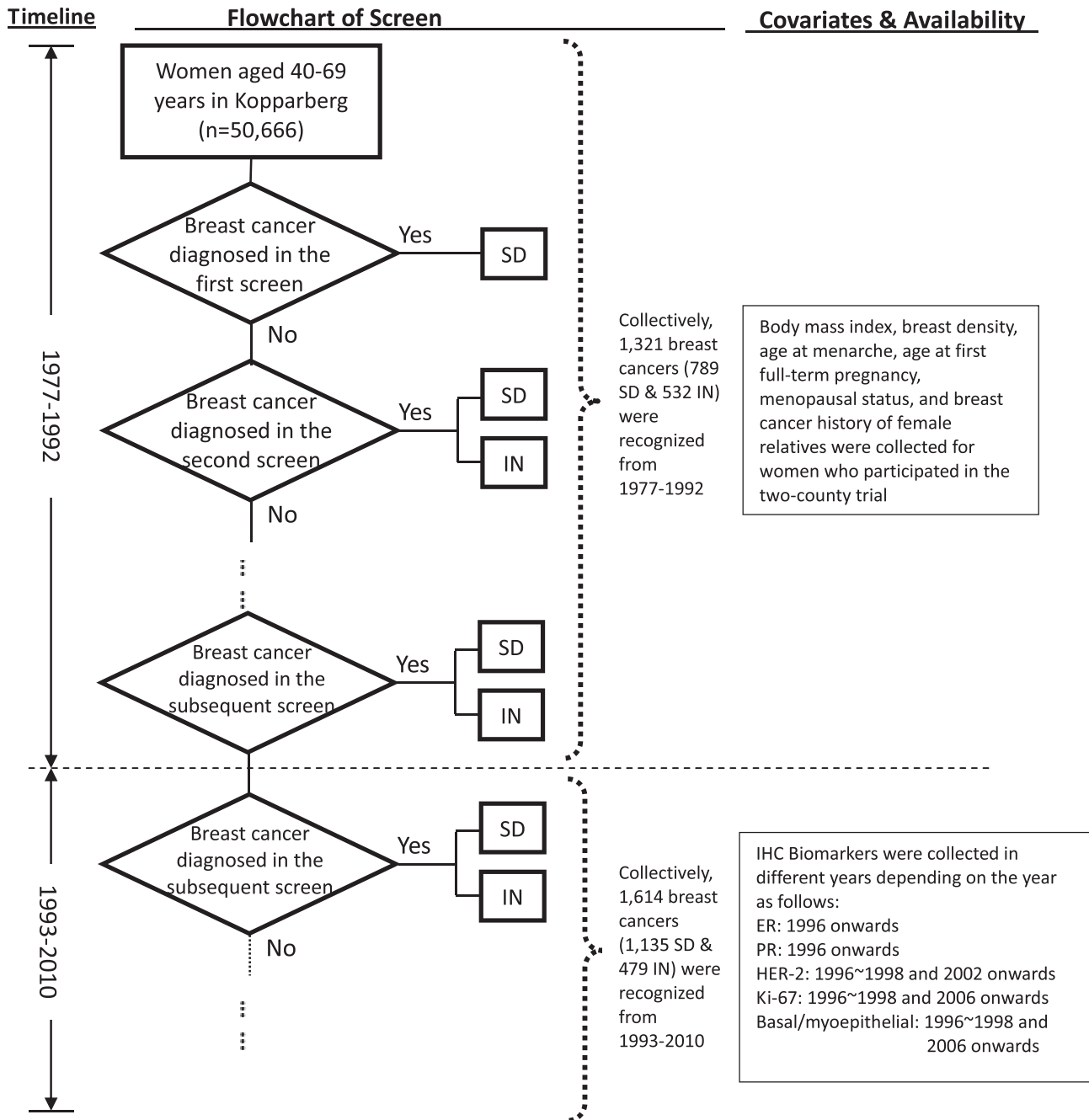


Fig. 2. The flowchart of the screen and data collection according to study period. SD, screen detected case; IN, interval case.

were responsible for the rate of entering the PCDP were classified as initiators, and those that were responsible for the progression from the PCDP to the CP were classified as promoters. We first considered the effect of single covariate simultaneously on $\lambda_1$ and $\lambda_2$ in the proposed four-state model. The model considering multiple covariates with forward selection was further applied. Statistical comparisons were two-sided, and a p-value of less than 0.05 was considered statistically significant.

## 3. Results

Women who were overweight/obese (BMI ≥25 kg/m$^2$), those with an early age at menarche, those with a late AP, and women

with a family history of breast cancer were more likely to enter the PCDP (eTable 2). However, the association was statistically significant ($P < 0.05$) only for the last two factors. Comparing screen-detected cases and interval cases, only overweight/obesity was statistically significant ($P = 0.005$). eTable 2 also shows the distributions of molecular markers in cases of breast cancer. ER and PR negativity, HER-2 and Ki-67 positivity, triple-negative status, and a basal-like phenotype were more frequently observed in clinically-detected breast cancers than in screen-detected breast cancers ($P < 0.05$).

With the application of a four-state continuous-time Markov model as shown in Fig. 3, the estimated annual incidence of PCDP breast cancer was 2.0 (95% CI, 1.97–2.12) per 1000 persons. The
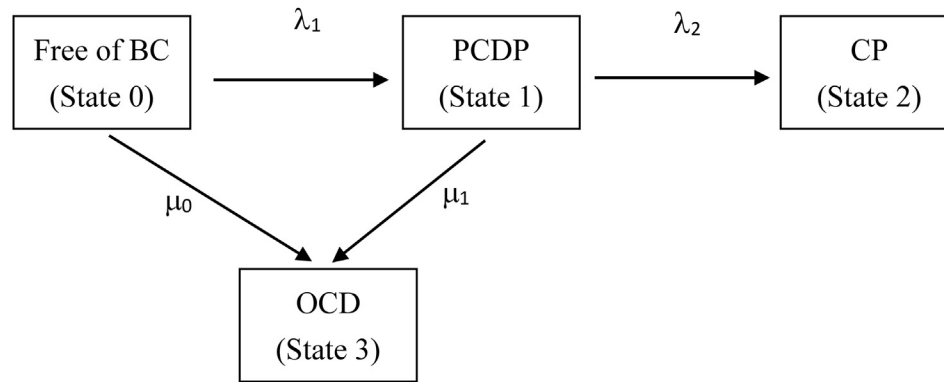
**Fig. 3.** The four-state Markov model for disease progression of breast cancer. BC, breast cancer; CP, clinical phase; OCD: other cause of death; PCDP, pre-clinical detectable phase. $\lambda_1$: the rate of entering the PCDP that would be affected by initiators $\lambda_2$: the progression rate from the PCDP to the CP that would be affected by promoters. $\mu_i$: the death rates due to causes other than breast cancer from State $i$ ($i = 0, 1$).

estimated transition rate from the PCDP to the CP was 0.38 (95% CI, 0.36–0.41), yielding 2.62 years (95% CI, 2.46–2.80) of MST. The death rates from other causes were 0.0062 (95% CI, 0.0061–0.0064) for women free of breast cancer (State 0), and 0.0116 (95% CI, 0.0100–0.0133) for breast cancer in the PCDP.

This model was further extended to an exponential regression Markov model to investigate the joint effects of the roles of each factor in terms of initiators and promoters (Table 1). The results show that, as an initiator, late AP led to a significant 23% increased risk (RR 1.23; 95% CI, 1.10–1.39) of entering the PCDP; as a promoter, the 15% decreased risk was not statistically significant (RR 0.85; 95% CI, 0.71–1.03). BMI $\geq 25$ kg/m$^2$ led to a statistically significant 42% decrease in the risk of breast cancer in terms of its role as a promoter (RR 0.58; 95% CI, 0.46–0.73) but had a marginally significant 10% elevated risk in its role as an initiator (RR 1.10; 95% CI, 0.95–1.27). Compared with non-dense breast tissue, dense breast tissue was associated with significantly increased risks of 21% (RR 1.21; 95% CI, 1.02–1.44) as an initiator and 78% (RR 1.78; 95% CI, 1.37–2.32) as a promoter. Family history also played a role as an initiator (RR 1.88; 95% CI, 1.34–2.64), but not as a promoter (RR 0.83; 95% CI, 0.50–1.38).

Table 1 also shows the RR of molecular biomarkers as they related to promoters. The most notable biomarker was basal-like phenotype (RR 3.85; 95% CI, 2.34–6.31), followed by triple-negative status (RR 3.28; 95% CI, 2.28–4.70). Other significant risk factors that acted as promoters included ER negativity (RR 2.55; 95% CI, 1.95–3.33), PR negativity (RR 1.74; 95% CI, 1.39–2.18), HER-2 positivity (RR 1.54; 95% CI, 1.09–2.18), and Ki-67 positivity (RR 2.36; 95% CI, 1.69–3.29). When these biomarkers were reclassified into different molecular subtypes, we found that compared with Luminal A, the basal-like subtype was the most significant risk subtype for the promotion of disease progression (RR 4.24; 95% CI, 2.56–7.01).

The adjustment of multiple covariates for each other, as shown in Table 2, demonstrates that late AP, dense breasts, and a family history of breast cancer enhance the risk of entering the PCDP by 23% (95% CI, 10%–38%), 41% (95% CI, 19%–68%), and 89% (95% CI, 36%–163%), respectively (Table 2). It is interesting that overweight/ obesity may act as an initiator (adjusted RR [aRR] 1.15; 95% CI, 0.99–1.33) but was inversely associated with the risk of acting as a promoter (aRR 0.65; 95% CI, 0.51–0.82). Dense breasts (aRR 1.46; 95% CI, 1.12–1.91), triple-negative status (aRR 2.07; 95% CI, 1.37–3.15), and Ki-67 positivity (aRR 1.66; 95% CI, 1.19–2.30) were

**Table 1**
The results of relative risks for risk factors as initiators and promoters when considered independently in the continuous-time exponential regression Markov model.

| Risk factors | Initiators | Promoters |
|---|---|---|
| | RR$_1$ (95% CI) | RR$_2$ (95% CI) |
| Age at menarche, $\geq 12$ years vs. $\leq 11$ years | 0.84 (0.58, 1.20) | 1.06 (0.57, 1.95) |
| Age at first full-term pregnancy, $\leq 25$ years vs. >25 years | 1.23 (1.10, 1.39) | 0.85 (0.71, 1.03) |
| BMI, $\geq 25$ kg/m$^2$ vs. <25 kg/m$^2$ | 1.10 (0.95, 1.27) | 0.58 (0.46, 0.73) |
| Breast density, Dense vs. Non-dense | 1.21 (1.02, 1.44) | 1.78 (1.37, 2.32) |
| Family history, Yes vs. No | 1.88 (1.34, 2.64) | 0.83 (0.50, 1.38) |
| ER, Neg. vs. Pos. | NA | 2.55 (1.95, 3.33) |
| PR, Neg. vs. Pos. | NA | 1.74 (1.39, 2.18) |
| HER-2, Pos. vs. Neg. | NA | 1.54 (1.09, 2.18) |
| Triple-negative, Yes vs. No | NA | 3.28 (2.28, 4.70) |
| Ki-67, Neg. vs. Pos. | NA | 2.36 (1.69, 3.29) |
| Basal-like phenotype, Yes vs. No | NA | 3.85 (2.34, 6.31) |
| Molecular subtypes | | |
|     Luminal A | NA | 1.00 |
|     Luminal B | NA | 1.74 (0.99, 3.05) |
|     HER2+ | NA | 1.51 (0.69, 3.28) |
|     Basal-like subtype | NA | 4.24 (2.56, 7.01) |
|     Triple-negative | NA | 2.06 (0.75, 5.69) |

BMI, body mass index; CI, confidence interval; NA, not applicable; RR, relative risk.
RR$_1$: Relative risk for initiators as it relates to the risk of the development of breast cancer.
RR$_2$: Relative risk for promoters as it relates to the risk of progression from the PCDP to the CP.

**Table 2**
The results of relative risks for risk factors as initiators and promoters when considered jointly in the continuous-time exponential regression Markov model.

| Parameters | Model 1 | Model 2 |
|---|---|---|
| | aRR (95% CI) | aRR (95% CI) |
| **Initiators** | | |
| BMI, $\geq 25$ vs. $<25$ kg/m$^2$ | 1.15 (0.99, 1.33) | 1.15 (0.99, 1.34) |
| Age at first full-term pregnancy, $>25$ y vs. $\leq 25$ y | 1.23 (1.10, 1.38) | 1.23 (1.10, 1.38) |
| Breast density, Dense vs. Non-dense | 1.41 (1.19, 1.69) | 1.41 (1.19, 1.68) |
| Family history, Yes vs. No | 1.89 (1.36, 2.63) | 1.89 (1.36, 2.63) |
| **Promoters** | | |
| BMI, $\geq 25$ vs. $<25$ kg/m$^2$ | 0.65 (0.51, 0.82) | 0.65 (0.52, 0.81) |
| Breast density, Dense vs. Non-dense | 1.46 (1.12, 1.91) | 1.46 (1.13, 1.89) |
| Triple-negative, Yes vs. No | 2.07 (1.37, 3.15) | NA |
| Ki-67 expression, Pos. vs. Neg. | 1.66 (1.19, 2.30) | NA |
| Basal-like phenotype, Yes vs. No | 1.71 (0.95, 3.10) | NA |
| **Molecular subtype** | | |
| Luminal A | NA | 1.00 |
| Luminal B | NA | 1.74 (0.99, 3.04) |
| HER-2+ | NA | 1.51 (0.70, 3.26) |
| Basal-like subtype | NA | 4.24 (2.56, 7.02) |
| Triple-negative | NA | 2.06 (0.75, 5.69) |

aRR, adjusted relative risk; BMI, body mass index; CI, confidence interval; NA, not applicable.
Model 1: The effects of biomarkers were treated as independent.
Model 2: The combination of the status of ER, PR, HER-2 and basal-like phenotype was used to classify the cancer into 5 subtypes.

still significant promoters after they were adjusted with respect to each other. The role of a basal-like phenotype after adjustment for these factors was only marginally significant. When the molecular biomarkers were considered collectively as one classification (Model 2), the basal-like subtype was the most influential subtype that acted as a promoter (aRR 4.24; 95% CI, 2.56−7.02) compared with Luminal A. The Luminal B, HER-2-positive, and triple-negative subtypes had similar estimated results as shown in Table 1.

The estimated incidence of preclinical breast cancer ($\hat{\lambda}_1$), the transition rate from the PCDP to the CP ($\hat{\lambda}_2$), and the estimated MSTs of all possible risk profiles based on the combination of five covariates (80 subgroups) are provided in eTable 3. The rate of entering the PCDP ranges from the lowest (0.001) for women with a BMI $<25$ kg/m$^2$, early AP, non-dense breasts, and for those without a family history of breast cancer to the highest (0.0036) for the opposite characteristics (BMI $\geq 25$ kg/m$^2$, late AP, dense breasts, and a family history of breast cancer). Given the postulate that a shorter MST and a higher preclinical incidence rate imply a higher risk for the progression of breast cancer, we used the ratio of the MST to the preclinical incidence rate (M/I$_{PCDP}$) as a surrogate indicator for risk stratification. The lower the M/I$_{PCDP}$ is, the higher the risk. If the first (Q$_1$ = 603) and third quartiles (Q$_3$ = 1695) of M/I$_{PCDP}$ are considered the cut-off points, we defined those with an M/I$_{PCDP}$ less than Q$_1$ as high risk, those with an M/I$_{PCDP}$ larger than Q$_3$ as low risk, and those with an M/I$_{PCDP}$ between Q$_1$ and Q$_3$ as intermediate risk.

Fig. 4 shows the transition probabilities of the different characteristics of three subjects who each represent different risk groups.

Subject A: BMI $\geq 25$ kg/m$^2$, AP $> 25$, dense breasts, family history of breast cancer, basal-like subtype (High risk, $\hat{\lambda}_1 = 0.0036$, MST = 0.86 years, M/I$_{PCDP}$ = 237).
Subject B: BMI $<25$ kg/m$^2$, AP $> 25$, dense breasts, no family history of breast cancer, HER-2+ molecular subtype (Intermediate risk, $\hat{\lambda}_1 = 0.0022$, MST = 2.28 years, M/I$_{PCDP}$ = 1022).
Subject C: BMI $<25$ kg/m$^2$, AP $\leq 25$, non-dense breasts, no family history of breast cancer, Luminal A molecular subtype (Low risk, $\hat{\lambda}_1 = 0.0010$, MST = 3.44 years, M/I$_{PCDP}$ = 3597).

The 2-year transition probabilities from "breast cancer-free" to the CP decreased from 0.44% for Subject A to 0.05% for Subject C. Fig. 4a and b show dynamic curves that indicate the evolution of breast cancer from "breast cancer-free" through the PCDP and finally to the CP following the temporal course of the natural history of breast cancer. A higher-risk subject, such as Subject A, had higher odds of progression from the PCDP to the CP (Fig. 4c), which led to a lower chance of entering the PCDP (Fig. 4b) following a fixed cohort study after 10 years of follow-up.

## 4. Discussion

We used cases of screen-detected breast cancer and cases of clinically-detected interval breast cancer obtained from a longitudinal breast cancer study in Sweden, in conjunction with a four-state continuous-time exponential regression Markov model, to assess the role of each risk factor played in both the initiation and the progression of breast cancer. We found that BMI, AP, breast density, and family history of breast cancer are regarded as initiators, while BMI, breast density, triple-negative status, Ki-67 expression, and basal-like phenotype or molecular subtypes are promoters. This method was very useful for the investigation of the impact of these risk factors on the occurrence of breast cancer (entering the PCDP), and also on the MST, which is the average time of progression from the PCDP to the CP.

BMI is a well-known risk factor for breast cancer: one meta-analysis reported that the risk ratio for breast cancer was 1.12 (95% CI, 1.08−1.16) for each 5 kg/m$^2$ increment.[21] In our study, the hazard ratio for the incidence of preclinical breast cancer for women with a BMI greater than 25 kg/m$^2$ was equal to 1.15 when other risk factors were taken into account. Furthermore, the hazard ratio for the transition from the PCDP to the CP for women with a BMI greater than 25 kg/m$^2$ was 0.65, which implies a slower transition toward the CP (analogous to a longer mean sojourn time in the PCDP) than women with a BMI less than 25 kg/m$^2$. The study by Kricker also found that a low BMI was highly correlated with clinically-detected cancers.[22] Subjects with a higher BMI had a higher risk of developing breast cancer, but the progression rate to the CP was relatively slow; this implies that those subjects can be easily identified in the screening program.
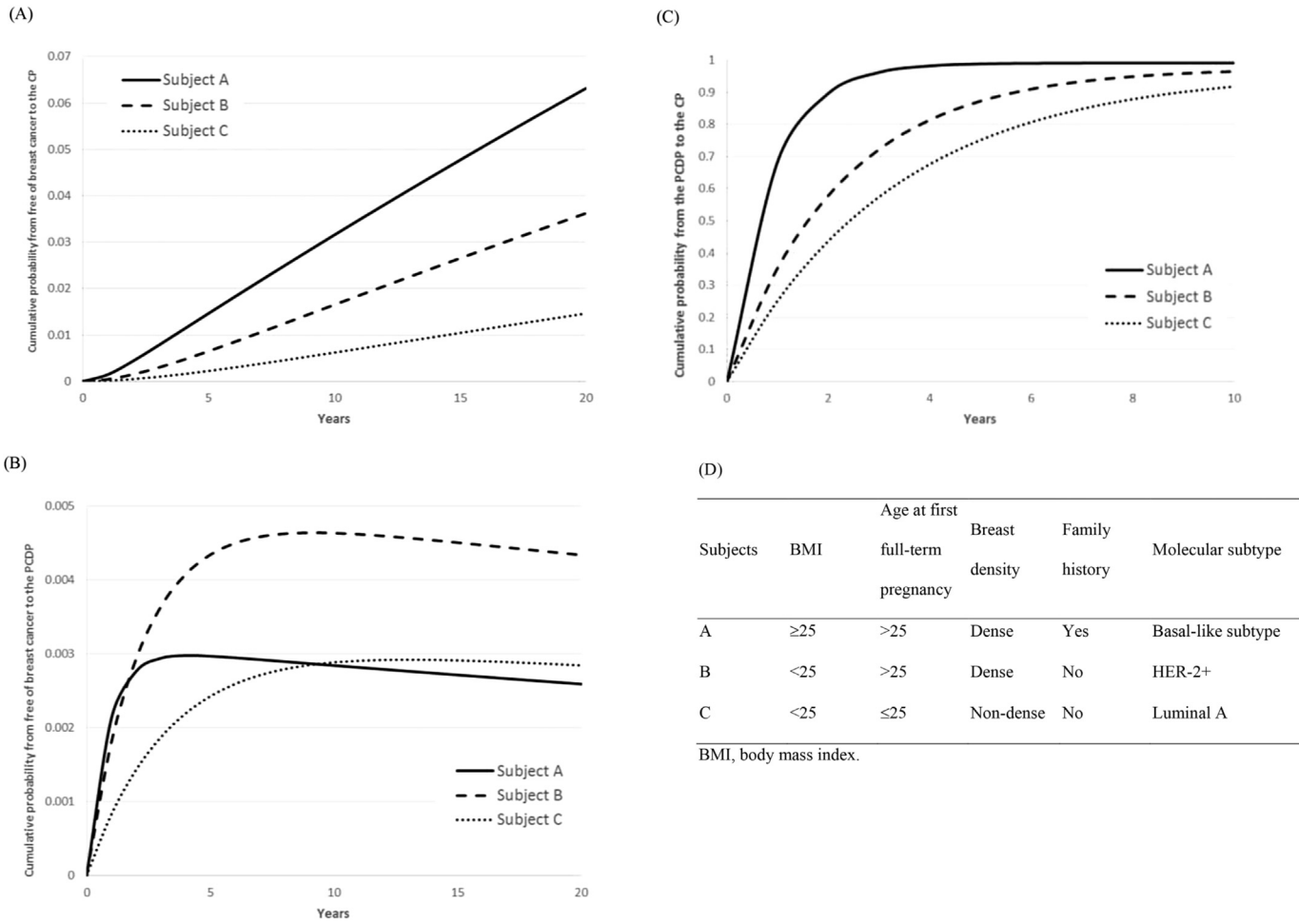
(A)



(B)



(C)



(D)

| Subjects | BMI | Age at first full-term pregnancy | Breast density | Family history | Molecular subtype |
|---|---|---|---|---|---|
| A | ≥25 | >25 | Dense | Yes | Basal-like subtype |
| B | <25 | >25 | Dense | No | HER-2+ |
| C | <25 | ≤25 | Non-dense | No | Luminal A |

BMI, body mass index.

**Fig. 4.** The illustration of the cumulative probabilities by time of three hypothetical subjects in different risk groups from high risk (Subject A) to low risk (Subject C).

Reproductive factors, including age at menarche, AP, age at menopause, and number of live births, are believed to exert protective effects due to the hormonal changes during pregnancy and lactation.[23] We did not find any significant reduction in risk in women with an older age at menarche in our study cohort, but women with an AP greater than 25 years had a higher risk of entering the PCDP. However, this factor did not act as a promoter. Moreover, the radiographic appearance of the breast was a consistent and strong risk factor for breast cancer.[24] The study by Chiu first described how to quantify the role of breast density on the preclinical incidence rate and the MST.[25] Based on the multistate model, the hazard ratio of the transition from "breast cancer-free" to the PCDP for women with dense versus non-dense breasts was 1.65, and that of the transition from the PCDP to the CP was 1.61; however, both estimates did not account for the adjustment of other covariates. After adjustments for BMI, AP, family history, and molecular type, the hazard ratios for dense breasts were 1.41 for the transition from "breast cancer-free" to the PCDP and 1.46 for the transition from the PCDP to the CP. The molecular subtypes of breast cancer play important roles in the prediction of the prognosis and are also considered when treatment modalities are chosen. The different molecular subtypes reflect various biologic characteristics of breast cancer tumors. Women with a shorter sojourn time may have the poorest survival. However, studies that have considered the MST according to different molecular types are lacking. Our estimate of the hazard ratio of the transition rate

from the PCDP to the CP corresponded to the estimate of the MST, of which the mean value was the inverse of the transition rate from the PCDP to the CP. A risk factor that acted as a promoter had a shorter MST. We found that the shortest MST was associated with a basal-like subtype, which is consistent with the finding that patients with basal-like breast cancer experience poorer survival.[8] Furthermore, a shorter MST indicates a smaller chance that the cancer will be detected by screening. In our study, 18% of interval cancers were the basal-like phenotype, whereas only 5% of screen-detected cases showed the basal-like phenotype.

The results of the current study can serve as a predictive model for the incidence of preclinical breast cancer and to determine the likelihood that preclinical breast cancer will transition to the CP. Predictive modeling of breast cancer has had a long history since 1983. The first was Pike's model, which predicted the risk of breast cancer with the evolution of hormonal risk factors, including age at menarche, AP, and age at menopause, together with weight.[1] The subsequent seminar work on the predictive model from Gail's study was developed to integrate these hormonal risk factors.[3] Several revisions and extensions, including the addition of postmenopausal hormone levels (such as those of estrogen), body weight, and alcohol consumption, were proposed by Colditz et al.[26] Another addition includes a consideration of breast density (Barlow et al.[4] introduced genetic markers from the Claus model[27] and estimated the carrier probability of a BRCA 1/2 mutation from Couch et al. and Parmingiani et al.[28,29]). Despite this abundance of

information, previous work did not consider the risk of breast cancer according to the preclinical and clinical phases, which has become essential due to the current worldwide popularity of screening. Nevertheless, the previous models suggest that, in our model, we can further treat genetic markers as initiators when the data become available.

### 4.1. Methodological consideration

In the current study, we investigated the joint effects of risk factors for initiators and promoters using a time-continuous four-state Markov model, which is different from the conventional methods. Using conventional analysis methods, we identified risk factors through a comparison of the distribution of risk factors between women free of breast cancer and women with breast cancer (for initiators). We also compared the distribution between screen-detected and interval cancers (for promoters), as shown in eTable 2. Although this is intuitive and the data are easy to access, such a method could only be used to conduct an initial exploration due to several shortcomings. First, for initiators, we had to compare women free of breast cancer with those who had been diagnosed with breast cancer, regardless of whether a particular woman had screen-detected breast cancer or interval breast cancer. However, the inclusion of interval cancer means that we have not only introduced the influence of disease occurrence but also the transition from the PCDP to the CP. The two entangled forces of interval cancers precluded the declaration that such a comparison was purely for initiators. Second, the conventional means by which two separate/independent comparisons are made could not capture the covariance between two transitions (from "breast cancer-free" to the PCDP and from the PCDP to the CP) and their regression coefficients. However, the biased standard error would affect the inference of statistical significance. Third, the conventional method failed to provide absolute estimates of the incidence and the MST via the risk profiles, which is important for individually tailored screening and also for surveillance programs.

Our four-state continuous-time exponential regression Markov model enabled us to investigate the effects of risk factors on the transition from "breast cancer-free" to the PCDP (initiators) and from the PCDP to the CP (promoters) using data on screen-detected and clinically-detected interval breast cancer. Although many studies have investigated the risk factors of breast cancer by conventional survival analysis, such as the Cox proportional hazards regression model, these methods can only identify stage-specific covariates with two separate models instead of the correlated transitions of two states between "free of breast cancer" and the PCDP and also between the PCDP and the CP. Moreover, even the two separate models with the conventional survival analysis required an explicit definition in terms of the start and end time to an event, which was impossible to observe for the transition from the PCDP to the CP in our case, as the PCDP is expressed as a screen-detected case. The further transition from the PCDP to the CP would be interrupted by medical intervention. With the Markov model, we were able to estimate this hidden transition rate because it was embedded in the second transition following the first transition from "free of breast cancer" to the PCDP among those with interval cancers.

### 4.2. Clinical implication

The elucidation of the respective contributions from relevant factors is valuable for the prevention of breast cancer. From a practical aspect, the identification of initiators can help in the design of individually tailored prevention programs, and the elucidation of promoters may enable us to consider individually tailored surveillance strategies. The logarithm of the transition rate from "breast cancer-free" to the PCDP can be taken as a risk score, which allows us to stratify the population into different risk categories for individualized prevention programs. Information on individual breast cancer risk (initiator) may be tailored to offer preventive interventions, such as diet, exercise, and hormonal therapy. For secondary prevention, women at a higher risk would be recommended to undergo screening at an earlier age.[30] Information on the MST can help to provide a shorter screening interval or a more intensive surveillance strategy for those with a higher risk (more promoters) to reduce false-negative and false-positive cases. Furthermore, factors that are unknown during screening, such as the molecular subtype, can be used to design individualized surveillance regimes for women once they have been diagnosed with cancer because the transition from the PCDP to the CP is an indicator of tumor aggression.

The ratio of MST to preclinical incidence rate ($M/I_{PCDP}$) can be used as an indicator of whether it is economical to screen women with a given risk based on the criteria that the optimal value would be ideal for a specific screening policy under consideration. The lower the $M/I_{PCDP}$, the more cost-effective it will be to screen women with a given risk. The values range from 176 to 4832. The median value is approximately 1,004, which represents the women with an average risk. If a triennial screening policy is applied to the women with an average risk, it is clear that a value greater than 3000 would not be cost-effective if the screening policy was based on an average-risk group. A screening program with a longer interval, such as 8 years, and less aggressive adjuvant therapy would be suggested. If the value was less than 500, the screening of such high-risk women would be very cost-effective if an annual screening program is chosen and more aggressive therapy is applied.

We ascertained state-dependent covariates of initiators and promoters to classify the risk of the two-step progression of breast cancer. The application of such a risk assessment model to the temporal and natural course of breast cancer contributes to population-based risk stratification of patients with breast cancer and also to our understanding of the subsequent progression to advanced breast cancer. Such information is very helpful for individually tailored screening programs and for personalized clinical surveillance of early-detected breast cancer.

### Funding

### Conflicts of interest

None declared.

### Author's contribution

Amy Ming-Fang Yen and Wendy Yi-Ying Wu were responsible for statistical analysis, interpretation of the results, and drafting of the manuscript. Laszlo Tabar was responsible for data collection, and interpretation of the results. Stephen Duffy and Rober Smith contributed to the interpretation of the results, and the revision of the manuscript. Hsiu-Hsi Chen contributed to concept formulation, study design, statistical analyses, interpretation of results, and the revision of the manuscript. All authors have read the manuscript and approve its submission.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.je.2016.10.003.

## References

1. Pike MC, Krailo MD, Henderson BE, et al. 'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer. *Nature*. 1983;303:767–770.
2. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol*. 1983;13:1141–1151.
3. 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;81:1879–1886.
4. Barlow WE, White E, Ballard-Barbash R, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. *J Natl Cancer Inst*. 2006;98:1204–1214.
5. Tice JA, Cummings SR, Smith-Bindman R, et al. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med*. 2008;148:337–347.
6. Amir E, Freedman OC, Seruga B, et al. Assessing women at high risk of breast cancer: a review of risk assessment models. *J Natl Cancer Inst*. 2010;102: 680–691.
7. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. *Breast Cancer Res Treat*. 2012;132:365–377.
8. Cheang MC, Voduc D, Bajdik C, et al. Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin Cancer Res*. 2008;14:1368–1376.
9. Dong W, Berry DA, Bevers TB, et al. Prognostic role of detection method and its relationship with tumor biomarkers in breast cancer: the university of Texas M.D. Anderson Cancer Center experience. *Cancer Epidemiol Biomark Prev*. 2008;17:1096–1103.
10. Hsieh HJ, Chen TH, Chang SH. Assessing chronic disease progression using non-homogeneous exponential regression Markov models: an illustration using a selective breast cancer screening in Taiwan. *Stat Med*. 2002;21:3369–3382.
11. Tabar L, Fagerberg CJ, Gad A, et al. Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the breast cancer screening working group of the Swedish national board of Health and Welfare. *Lancet*. 1985;1:829–832.
12. Tabar L, Gad A. Screening for breast cancer: the Swedish trial. *Radiology*. 1981;138:219–222.
13. Gram IT, Funkhouser E, Tabar L. The Tabar classification of mammographic parenchymal patterns. *Eur J Radiol*. 1997;24:131–136.
14. Wolfe JN. Breast parenchymal patterns and their changes with age. *Radiology*. 1976;121:545–552.
15. Tot T, Pekar G. Multifocality in "basal-like" breast carcinomas and its influence on lymph node status. *Ann Surg Oncol*. 2011;18:1671–1677.
16. Duffy SW, Chen HH, Tabar L, et al. Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. *Stat Med*. 1995;14:1531–1543.
17. Chen HH, Duffy SW, Day NE. Markov chain models for progression of breast cancer. Part II: prediction of outcomes for different screening regimes. *J Epidemiol Biostat*. 1997;2:23–35.
18. Chen HH, Duffy SW, Day NE. Markov chain models for progression of breast cancer. Part I: tumour attributes and the preclinical screen-detectable phase. *J Epidemiol Biostat*. 1997;2:9–23.
19. Chen TH, Kuo HS, Yen MF, et al. Estimation of sojourn time in chronic disease screening without data on interval cases. *Biometrics*. 2000;56:167–172.
20. Taghipour S, Banjevic D, Miller AB, Montgomery N, Jardine AK, Harvey BJ. Parameter estimates for invasive breast cancer progression in the Canadian national breast screening study. *Br J Cancer*. 2013;108:542–548.
21. Renehan AG, Tyson M, Egger M, et al. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet*. 2008;371:569–578.
22. Kricker A, Disipio T, Stone J, et al. Bodyweight and other correlates of symptom-detected breast cancers in a population offered screening. *Cancer Causes Control*. 2012;23:89–102.
23. Thomas DB. Do hormones cause breast cancer? *Cancer*. 1984;53:595–604.
24. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med*. 2007;356:227–236.
25. Chiu SY, Duffy S, Yen AM, et al. Effect of baseline breast density on breast cancer incidence, stage, mortality, and screening parameters: 25-year follow-up of a Swedish mammographic screening. *Cancer Epidemiol Biomark Prev*. 2010;19:1219–1228.
26. Colditz GA, Rosner B. Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. *Am J Epidemiol*. 2000;152:950–964.
27. Claus EB, Risch N, Thompson WD. Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction. *Cancer*. 1994;73:643–651.
28. Couch FJ, DeShano ML, Blackwood MA, et al. BRCA1 mutations in women attending clinics that evaluate the risk of breast cancer. *N Engl J Med*. 1997;336: 1409–1415.
29. Parmigiani G, Berry D, Aguilar O. Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. *Am J Hum Genet*. 1998;62: 145–158.
30. Wu YY, Yen MF, Yu CP, et al. Individually tailored screening of breast cancer with genes, tumour phenotypes, clinical attributes, and conventional risk factors. *Br J Cancer*. 2013;108:2241–2249.