Neoplasma 2022; 69(3): 708–722

doi:10.4149/neo_2022_211118N1642

Development of a model for predicting risk from breast interval cancer in the female population in the Republic of Croatia

Romana Tandara HAČEK[1,*], Nataša ANTOLJAK[2], Marijan ERCEG[2]

[1]Health Care Quality Service, Ministry of Health, Zagreb, Croatia; [2]Division for Epidemiology and Prevention of Noncommunicable Chronic Diseases, Croatian Institute of Public Health, Zagreb, Croatia

**Correspondence: rhacek99@gmail.com*

Received November 18, 2021 / Accepted February 15, 2022

There are several risk prediction models for screen-detected breast cancer but to the best of our knowledge, none for predicting risk from the interval cancer in breast cancer screening. The challenge for developing such a model was that the risk factors for both cancers appear to be similar, but the effects of interval cancer on women's health are more severe due to its higher biological aggressiveness. Our model is based on risk factors identified in the female population in the Republic of Croatia. Anonymized data from 472,395 women who participated in the National Program for Early Detection of Breast Cancer during the first three cycles of the program (October 2006-May 2014) were used. Cancer data from the Breast Cancer Screening Registry were linked by the data linkage method with data from the Cancer Registry of the Republic of Croatia. A total of 789 women with interval cancer and 3,530 women with screen-detected cancer were identified. Multivariate logistic regression in R was used to model the difference between participants with screen-detected cancer and those with interval cancer, using the general linear model (glm) function. The variables used for the analysis were selected using the all subset regression analysis method. The criterion of the least complexity parameter, the Cp-Mallows index, was chosen. Three variables were found to be statistically significant in the model: breast tissue density ($p=0.038$), hormone replacement therapy ($p=0.034$), and a first-degree family history of breast cancer ($p<0.001$). The resulting model has a discriminant accuracy of 0.658 (95% CI 0.602–0.713). Although our model has poorer predictor reliability, its advantage is that it is based on real-world data and that the criteria for interval cancer were strictly followed. It is best suited for use in the Croatian population of women because we have identified the available risk factors for the development of interval cancer in our population, but with knowledge of a specific epidemiological environment, it can be more widely applied. The model can be used to make recommendations for individual screening participants. The variables of breast tissue density and first-degree family history of breast cancer increase the likelihood of interval cancer and indicate an increased risk of detecting interval cancer between mammograms. Consequently, individualized risk screening should be considered (modification of screening interval or additional screening by magnetic resonance or ultrasound). According to the model, hormone replacement therapy is positively related to screen-detected cancer, and participants who use hormone replacement therapy must be medically monitored due to the increased risk of screen-detected cancer. In addition, participants in the screening program who use hormone replacement therapy and have a higher density of breast tissue should be encouraged to have more frequent mammograms.

Key words: cancer breast screening; interval breast cancer; risk prediction model

In Croatia, the National Program for Early Detection of Breast Cancer has been implemented since 2006 using the mammography screening method. The program covers women aged 50–69 who are referred for mammography every two years. The Ministry of Health is the holder of the program and the Croatian Institute of Public Health (CIPH) plans, organizes, implements, and coordinates the program. The organization and quality assurance of mammography screening is carried out in accordance with the European Guidelines for Quality Assurance of Screening and Diagnosis of Breast Cancer [1].

In order to obtain initial evidence of the effectiveness of breast cancer screening programs, even before the expected long-term reduction in breast cancer mortality rates occurs, the European guidelines recommend the assessment and evaluation of the so-called surrogate indicators, which include the rate of interval cancers and their stage. According to the above guidelines, interval cancer is primary breast cancer diagnosed in a woman who has participated in screening with or without further assessment and which was negative for malignancy before the next screening invitation or within a period corresponding to the screening interval

if the woman has reached the upper age limit for participation in screening [1]. At the CIPH, there is a Cancer Registry for the Republic of Croatia. It is possible to obtain data on interval cancer and evaluate the impact of the program by linking its data with the Register of Deceased Persons. The goal of interval cancer surveillance is twofold, as a radiological revision of mammograms of women diagnosed with interval cancer serves for quality assurance and training of program providers [2–5]. The completeness of cancer data collected in screening registries is particularly important for the detection of interval cancer. The comparability of interval cancer between different populations may be limited by incomplete data in the application [6].

Interval breast cancer characteristics. It is known that interval breast cancer may have different characteristics than screen-detected cancer. This cancer can often be more aggressive, has a higher histological grade, larger tumor size, and a higher TNM stage (primary tumor (T sign), condition of regional lymph nodes (N sign), presence of distant metastases (M sign)). Heidinger et al. (2012) found that the percentage of T2 to T4 malignancies was significantly higher in women with interval cancer than in those with screen-detected cancer [6]. Bellio et al. (2017) showed that interval cancer has more aggressive characteristics compared to cancer detected by the screening, such as tumor invasiveness, tumor size, stage, and St. Gallen molecular subtype, resulting in a higher metastasis rate, worse overall survival, and worse disease-free survival [7]. Meshkat et al. (2015) discovered that interval breast cancer was less likely to be positive for the estrogen receptor and significantly more likely to overexpress human epidermal growth factor receptor 2 (HER2) than screen-detected cancer [8]. According to Defossez et al. (2018), interval cancer was diagnosed at a later stage than screen-detected cancer, but with significantly fewer metastases than cancer detected outside of the screening program [9]. The association between ductal cancer in situ (DCIS) and subsequent invasive interval cancer was studied by Duffy et al. (2016). DCIS detected at the screening was found to have a significant negative association with the rate of invasive interval cancer, implying that DCIS detection and treatment are critical for preventing future invasive cancer [10].

Risk factors for interval breast cancer. The risk factors for developing interval cancer have not been fully explored [8, 10–15].

According to the literature, approximately 13% to 19% of women diagnosed with breast cancer had first-degree relatives (mother, daughter, or sister) compared to women without breast cancer (8–12%) [16, 17]. The risk of developing breast cancer increases 1.5 to 4 times with the number of first-degree relatives [16, 18, 19]. Nonetheless, the results of previous publications on the association of family history and interval breast cancer are contradictory. Holm et al. (2015) believe that the reason lies in the use of different definitions of family history and the small number of subjects with interval breast cancer in the studies [12]. In the research conducted by Roman et al. (2017), women with a family history of breast cancer had an increased cumulative risk of both screen-detected and interval cancer [20]. True interval cancer was most strongly linked to a family history of breast cancer, according to Blanch et al. (2014) [11].

A woman's youthful age, which partially reflects breast density, is cited by Houssami and Hunter (2017) as a confounding factor in risk computation [21]. Contrary to the claim that higher breast density is a risk factor for developing interval cancer, Holm et al. (2015) found that interval cancer in women with lower breast density has a more aggressive phenotype. Furthermore, interval cancer in non-dense breasts (20% mammographic density) was significantly more likely to have lymph node involvement than screen-detected breast cancer, whereas interval cancer in dense breasts (>40.9% mammographic density) was less aggressive and phenotypically more similar to screen-detected breast cancer [12].

According to Evans and Howell (2015), weight gain prior to menopause, as well as being overweight or obese throughout menopause, increases the risk of breast cancer [22]. Strand et al. found that high BMI was linked to the likelihood of discovering a tumor larger than 2 cm at diagnosis. Women with interval cancer who have a high BMI have a worse prognosis, and they should be encouraged to participate in screening [23]. On the other hand, Boyd et al. (2014) highlighted low body mass index (BMI) as a factor associated with a higher relative incidence of interval cancer [24]. A negative relationship between interval cancer and a BMI greater than 25 kg/m^2 was also discovered by Holm et al. (2015) [12].

Hsieh et al. proved in 1990 that early onset of menarche and late menopause are factors that increase breast cancer risk and that postponing menarche by two years reduces breast cancer risk by about 10% [25]. Kelsey et al. (1993) discovered that the risk of breast cancer increases by about 17% every 5 years of menopausal age [26]. In postmenopausal women, the higher risk associated with older age at natural menopause generally does not occur before the age of 65, implying that the effect of menopausal age is not visible 10–20 years after menopause [26]. The Oxford research group found that each year of earlier menarche raises the risk of breast cancer significantly more than each year of later menopause. They concluded that menarche and menopause affect breast cancer risk not only by extending a woman's overall reproductive age but also that endogenous ovarian hormones are much more important for estrogen receptor-positive disease than for receptor-negative disease, as well as for lobular than for ductal tumors [27].

During pregnancy, hormones affect metabolism, gene expression, and mammary epithelial cell (MEC) proliferation dynamics. Pregnancy before the age of 20 reduces the probability of developing breast cancer by 50% compared to being nulliparous. Women who have their first pregnancy between the ages of 30 and 34 have no protective effect of pregnancy, whereas women who have their first pregnancy

after the age of 35 have an increased risk of breast cancer (according to Merrill et al., 5%) [28–30]. The overall risk of developing breast cancer increases immediately after delivery and is independent of race, age, or the number of pregnancies. Callihan et al. (2013) showed that, in comparison to nullipara, patients diagnosed with breast cancer in the first 5 years following pregnancy had a 2.8-fold higher risk of metastases and a 2.7-fold higher risk of mortality [31]. According to Slepicka et al. (2019), the impact of hormonal fluctuations, as well as the mammalian parous extracellular matrix and immune compartments, on tumorigenic potential modulation and the persistence of parity-induced transcription in mammary epithelial cells, should be considered [32].

Women who were currently using hormonal contraception or had recently used hormonal contraception (within the last 6 months) had a higher relative risk of breast cancer than women who had never used hormonal contraception, according to Mørch et al. (2017). The risk of breast cancer increased with the duration of use, and women who used hormonal contraceptives for more than 5 years had an increased risk for at least 5 years after stopping therapy. Gaffield et al. conducted a systematic review of the MEDLINE and CENTRAL databases (1966–2008) to investigate the effects of oral contraception on breast cancer risk in women with a positive family history of breast cancer. The majority of studies (10 studies and a pooled analysis of 54 studies) indicated that oral contraception did not increase risk in these women. Only four studies indicated an increased risk of breast cancer, particularly in women who used oral contraceptives prior to 1975 [33]. According to the World Health Organization (2015), hormonal contraception does not increase the risk of breast cancer in a population of women with a family history of breast cancer (first- or second-generation relatives) and should not be limited in its use [34]. We concluded from the literature [28–40], that the effect of hormonal contraception on breast cancer risk is contradictory and dependent on the variants of hormones in their composition.

The Oxford Collaborative Group on Hormonal Factors in Breast Cancer (2019) conducted a meta-analysis of data on the use of hormone replacement therapy (HRT), (1992–2018). Each type of HRT, except vaginal estrogens, was associated with excess breast cancer risk, which increased steadily with the duration of use and was higher for the estrogen-progestin combination compared to estrogen alone. Some excess risks persist ten years after HRT discontinuation [27].

This paper aims to present a model for predicting interval breast cancer based on risk factors identified in the female population in the Republic of Croatia.

Patients and methods

Anonymized data from the screening database of women aged 50 to 69 who participated in mammography during the first three cycles of the National Program for Early Detection of Breast Cancer in the Republic of Croatia (October 2006–May 2014) were used. The invitation surveys were fulfilled by women prior to screening and served as the source of data. The findings of mammography readings according to the classification BI-RADS performed by two independent certified radiologists were entered into the screening register. Data on cancer were obtained from onco-leaflets and reports of malignant neoplasms as well as from the Cancer Registry. The screening database contains data on 472,395 women who participated in the screening.

The use of data from the screening database and the Cancer Registry of the Republic of Croatia required prior approval from the Ministry of Health of the Republic of Croatia and the CIPH Ethics Committee.

Cancer data from the screening registry from the first to the third cycle of the first screening round were linked in CIPH using the method of data linkage to the data from the Cancer Registry of the Republic of Croatia.

The first screening cycle included women between the ages of 50 and 72, while the second and third included women between the ages of 50 and 71. Interval cancer was defined as BI-RADS 1 or 2 cancer with a time interval of 365 to 791 days between the date of the last mammogram and the date of registration in the Cancer Registry. Screen-detected cancer was defined as BI-RADS 0, 3, 4, or 5 cancers, with a time interval <366 days after mammography, and women whose date of registration in the Cancer Registry preceded the date of the most recent mammogram were excluded.

Results

Characteristics of the women in the first three cycles of the National Program for Early Detection of Breast Cancer in the Republic of Croatia (October 2006–May 2014) are presented in Table 1.

A total of 789 women with interval cancer and 3,530 women with screen-detected cancer were identified.

The incidence of interval cancer was 14.90% (n=286) in the first screening cycle, 20.35% (n=210) in the second cycle, and 21.42% (n=293) in the third screening cycle. The difference in the frequency of screen-detected versus interval cancer was statistically significant in relation to screening cycles (χ^2=26.633, number of degrees of freedom = 2, p<0.001).

Stages 0–I accounted for 58.45 percent (n=384) of all interval cancers, Stages IIA–IIIB for 37.90 percent (n=249), and Stage IV for 3.65 percent (n=24).

For 132 intervals and 742 screen-detected cancers, stage information was missing (Table 2).

Stage 0–I was the most common in screen-detected cancer, accounting for 65.03% (n=1,813), same as was for interval cancer, accounting for 58.45% (n=384).

The incidence of both cancers was lowest in Stage IV: 3.65% (n=24) for interval cancer and 1.54% (n=43) for screen-detected cancer.

Screen-detected cancer was more common than interval cancer in Stage 0–I (65.03% vs. 58.45%), while interval cancer

was more common in Stage IIA–IIIC (37.90% vs. 33.43%) and in Stage IV 3.65% vs. 1.54%).

The difference in cancer stage and cancer type frequency was statistically significant (χ^2=18.890, number of degrees of freedom = 2, p<0.001).

Since the average age at menarche in screening participants was 13.8±1.65 years, and the average age at menopause was 49.9±3.64 years, according to the preceding studies [25, 26, 41–46], only a subset of our screening participants would be at increased risk of breast cancer due to premature onset of menarche caused by prolonged exposure to endogenous hormones. The mean age of women in our screening was 60.7±5.99 years, with a range of 54.71 to 66.69 years, implying that most women in menopause and postmenopause participated in our screening. Natural menopausal women over the age of 65 may be at a slightly higher risk.

The majority of women in our screening (70.39%) used hormonal contraception for 1 to 5 years. According to Mørch et al. (2017), they have an increased risk for breast cancer at least 5 years after discontinuation [37]. Although we did not specify the type of hormonal contraception used in our study, it is reasonable to assume that the women who were screened after the first birth control pill was approved in 1960 also used the first generations of the pill [38]. Gaffield et al. presented evidence from four studies indicating an increased risk of breast cancer, particularly in women who used oral contraceptives prior to 1975. Therefore, we can assume that the sexually active generations of women in our screening, e.g., from 1960 to 1975, were at an increased risk of breast cancer [33]. Furthermore, oral contraceptives used during this time period were more strongly associated with breast cancer mortality, according to Charlton et al. [47].

Hormone replacement therapy (HRT) was used by 3.75% of participants, with an average age of 6.6 ± 3.85 years.

Relationship between risk factors (categorical variables) and interval cancer. The Chi-square test indicated that there was a statistically significant relationship between hormonal contraception and cancer (χ^2=4.21, number of degrees of freedom = 1, p=0.040), hormone replacement therapy and cancer (χ^2=6.98, number of degrees of freedom = 1, p=0.008) and breast density and cancer (χ^2=19.20, number of degrees of freedom = 2, p<0.001). It was not determined for first-degree family history of breast cancer and cancer (χ^2=2.65, number of degrees of freedom = 2, p=0.266) (Table 3).

Mann-Whitney U test. Mann-Whitney U test determined whether there was a statistically significant difference between the independent variables and dependent dichotomous categorical variable (interval cancer and screen-detected cancer). The difference between screen-detected and interval cancer was found to be statistically significant using the Mann-Whitney U test for the variable's height (Z=2.440, p=0.015), body weight (Z=–2.074, p=0.038), and year of first pregnancy (Z=3.359, p=0.001) (Table 4).

Univariate logistic regression. It was carried out for each of the independent variables to determine their correlation and predictive value. The R statistical program was used to analyze the results.

The dependent variable was the binary classification of women who had participated in screening into those with interval cancer (Ca-Int) and those with screen-detected cancer (Ca-Scr). For statistical purposes, interval cancer was coded as zero (0) and screen-detected cancer as one (1).

Independent variables were used based on the literature review and the results of the previous descriptive statistics and Chi-square test (Table 5). p-values for χ^2-tests deter-

Table 1. The characteristics of the women who participated in the screening.

Screening participants	Average value
Age (year)	60.7±5.99
50–54 (19.83%)	
55–59 (26.41%)	
60–64 (20.6%)	
65–69 (29.24%)	
70–74 (3.83%)	
Height (cm)	163.7±6.25
Weight (kg)	75.3±12.9
BMI	27.9(overweight)
Menarche (year)	13.8±1.65
Menopause (year)	49.9±3.64
Pregnancies (No.)	2.7±1.41
First pregnancy (year)	22.7±4.01
Hormonal contraception duration of use (year)	5.3±4.97
(21.15% of the participants)	
1–5 years (70.39%)	
6–10 years (19.72%)	
Hormone replacement therapy duration of use (year)	6.6±3.85
(3.75% the participants)	

Table 2. Stages of interval cancer identified in the first three cycles of the screening.

Type of cancer	Stage 0–I Carcinoma *in situ* and localized carcinoma		Stage IIA–IIIC Regional lymph node metastases		Stage IV Distant metastases		Total	Missing data on Ca stage
	No.	%	No.	%	No.	%	No.	No.
Interval cancer (Ca-Int)	384	58.45	249	37.90	24	3.65	657	132
Screening cancer (Ca-Scr)	1813	65.03	932	33.43	43	1.54	2788	742
Total	2197		1181		67		3445	874

Table 3. The relationship between categorical variables and interval cancer using the χ^2-test test.

Variable	χ^2	Number of degrees of freedom	p-value
Hormonal contraception	4,21	1	0.040
Hormone replacement therapy	6.98	1	0.008
Breast density	19.20	2	<0.001
First-degree family history of breast cancer	2.65	2	0.266

Table 4. Mann-Whitney U test (Ca-Scr vs. Ca-Int).

Variables	Median	IQ range	Z adjusted	p-value
Height	164.0	160.0–168.0	**2.440**	**0.015**
Body weight	75.3	66.0–83.0	**–2.074**	**0.0380**
The year of the onset of menarche	13.8	13.0–15.0	–1.054	0.292
The year of the onset of menopause	49.9	48.0–52.0	0.610	0.542
Number of children	2.0	2.0–2.0	–1.489	0.1362
Number of pregnancies	2.0	2.0–3.0	0.306	0.7592
Year of the first pregnancy	22.0	20.0–25.0	**3.359**	**0.001**
Years of hormonal contraceptive use	4.0	2.0–7.0	–0.952	0.341
Years of hormone replacement therapy use	6.0	3.0–9.0	–0.940	0.3470
Age (years) at the time of diagnosis	56	47.0–65.0	0.0317	0.975

mined by statistical significance of variables at the 0.05 level. The 95% confidence limits for the interceptor and the coefficients of the variables used are also provided.

Four variables (shaded gray): breast density, year of first pregnancy, hormonal contraception, and hormone replacement therapy were statistically significant at the 0.05 significance level.

Multivariate logistic regression. The variables used for the analysis were selected using the all-subset regression analysis method. The criterion of the parameter of least complexity, Cp-Mallow's index, was chosen.

The key is in the Cp/p ratio and in reducing the number of variables in the model while maintaining the same predictive power [48–50].

The glm (general linear model) function in R was used to model multivariate logistic regression. The dependent variable was a binary classification of participants into those with interval cancer, Ca-Int (0), and those with screen-detected cancer, Ca-Scr (1). In univariate logistics regression, statistically significant variables at the 0.05 level were used as independent variables. Two logistic regression variants were tested in modeling: stepwise forward regression and stepwise backward regression [51–53]. Because highly pre-selected variables were included, both logistic regression variants produced identical results.

The resulting model is shown in Table 6, and the odds ratios for individual variables are shown graphically (Figure 1) to make interpretation easier.

The distribution of the findings indicates that there are no cancer participants who deviate significantly from the model (maximum value 1.92). The coefficients for the different model components (interceptor + variable coefficients) are shown. The p-values for each model component are provided (Table 6).

Table 6 shows that three variables were found to be statistically significant in the model ($p<0.05$). These are breast tissue density ($p=0.038$), hormone replacement therapy ($p=0.034$), and a first-degree family history of breast cancer ($p<0.001$). The variables of breast tissue density and first-degree family history of breast cancer increase the likelihood of interval cancer, whereas hormone replacement therapy is positively related to screen-detected cancer (Figure 1).

The prediction model has an area under the ROC curve of 0.658, with 95% confidence limit (CI) ranging from 0.602 to 0.713 (Figure 2).

As the area under the ROC curve is a measure of the accuracy of the test, a test with CI between 0.60–0.70 should be considered as a weaker test.

Discussion

To the best of our knowledge, there are several models for predicting screen-detected breast cancer, but none for predicting risk from interval breast cancer in breast cancer screening.

Nguyen et al. (2020) created an interval breast cancer risk model. They claim that no valid risk model exists to predict interval breast cancers. Within the Melbourne Collaborative Cohort Study, they conducted a nested case-control study with 168 interval breast cancer patients and 498 matched control subjects. Body mass index (BMI) and first-degree family history obtained via a questionnaire were included in the model. Instead of BI-RADS tools, they used the CUMULUS software to measure breast density and age-adjusted breast tissue aging, a novel measure of estrogen and progesterone exposure. They fitted conditional logistic regression to estimate the odds ratio (OR) or odds ratio per adjusted standard deviation (OPERA) and calculated the area under the receiver operating characteristic curve (AUC). They concluded that compared with using dense breasts alone, risk discrimination for interval breast cancers could be doubled by using breast density, BMI, family history, and hormonal exposure. Their model improves risk predictions and clinical recommendations for breast screening procedures, alerting women with dense breasts and their doctors to an increased risk of developing interval breast cancer. Rosner et al. validated the model and extended it including one or more full-term births [54, 55].

The challenge for developing our model was that the risk factors for both cancers appear to be similar, but the conse-

Table 5. The results of the of the univariate logistic model analysis.

1. First-degree of a family history of breast cancer		χ^2=2.004769	df=1	p=0.1568149	
		Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
	(Intercept)	170.000	118.000	245.00	3.82e-168
	Heredity	0.644	0.525	0.79	2.44e-05
2. Density		χ^2=18.89885	df =1	p=0.0000138	
		Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
	(Intercept)	7.050	5.670	8.760	3.38e-69
	Density	0.778	0.694	0.871	0.8711.27e-05
3. Height		χ^2=2.456791	df=1	p=0.1170275	
		Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
	(Intercept)	12.500	3.220	48.8	0.000266
	Height	0.994	0.986	1.0	0.141000
4. Body mass		χ^2=2.305832	df=1	p=0.1288991	
		Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
	(Intercept)	3.17	1.980	5.06	0.00000142
	Body Mass	1.00	0.999	1.01	0.12700000
5. The year of the onset of menarche		χ^2=0.4723613	df=1	p=0.4919078	
		Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
	(Intercept)	3.96	2.530	6.19	0.00000000169
	Menarche	1.01	0.978	1.04	0.53700000000
6. The year of the onset of menopause		χ^2=0.5585541	df=1	p=0.4548489	
		Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
	(Intercept)	6.450	2.600	16.00	0.0000577
	Menopause	0.993	0.975	1.01	0.4600000
7. No. of children		χ^2=1.751093	df=1	p=0.1857489	
		Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
	(Intercept)	3.94	3.22	4.81	4.81 6.97e-41
	No Children	1.06	0.97	1.16	1.95e-01
8. No. of pregnancies		χ^2=0.0616637	df=1	p=0.8038870	
		Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
	(Intercept)	4.43	3.760	5.21	2.89e-72
	No Pregnancies	1.01	0.955	1.06	8.07e-01
9. Year of the first pregnancy		χ^2=8.827105	df=1	p=0.0029701	
		Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
	(Intercept)	8.200	5.370	12.500	2.07e-22
	Year First Pregn	0.974	0.957	0.991	3.59e-03
10. Hormonal contraception (HC)		χ^2=4.122075	df=1	p=0.0423354	
		Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
	(Intercept)	4.760	4.340	5.230	1.88e-238
	HC	0.828	0.692	0.992	4.05e-02
11. Years of HC usage		χ^2=0.5192016	df=1	p=0.4711877	
		Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
	(Intercept)	3.82	2.940	4.95	4.33e-24
	Years HC	1.01	0.977	1.05	4.74e-01

Table 5. ***Continued ...***

12. Hormone replacement therapy (HRT)		χ^2=6.370310	df=1	p=0.0116093	
		Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
	(Intercept)	4.64	4.28	5.04	9.88e-297
	HRT	1.61	1.13	2.30	8.83e-03
13. Years of HRT usage		χ^2=0.8576716	df=1	p=0.3543981	
		Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
	(Intercept)	2.11	1.140	3.89	0.0172
	Years HRT	1.0	0.958	1.12	0.3660
14. Age at time of Ca diagnosis		χ^2=0.0172364	df=1	p= 0.8955486	
		Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
	(Intercept)	5.150	1.64	16.20	0.00504
	Diagnosis Age	0.999	0.98	1.02	0.89500

Table 6. The results of the logistic model of discrimination between screen-detected and interval cancer.

Deviance Residuals:				
Min	**1Q**	**Median**	**3Q**	**Max**
−1.8382	−1.2188	0.7237	1.0553	1.9230
Coefficients:				
	Estimate	Std. Error	z value	Pr (> |z|)
(Intercept)	3.49941	0.78948	4.433	0.00000931***
Contraception	−0.42697	0.25588	−1.669	0.095198
Density	−0.34525	0.16599	−2.080	0.037531*
HormoneTherapy	1.84035	0.86649	2.124	0.033678*
YearFirstPregnan	−0.03291	0.02520	−1.306	0.191585
Heredity	−0.94467	0.25899	−3.647	0.000265***
Signif. codes:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			
Dispersion parameter for binomial family taken to be 1)				
Null deviance: 512.06 on 376 degrees of freedom				
Residual deviance: 480.46 on 371 degrees of freedom				
AIC: 492.46				
Number of Fisher Scoring iterations: 4				
	Odds ratio	**Lower 95% CI**	**Upper 95% CI**	**p-value**
(Intercept)	33.100	7.040	156.000	0.00000931
Density	0.708	0.511	0.980	0.03750000
HormoneTherapy	6.300	1.150	34.400	0.03370000
YearFirstPregnan	0.968	0.921	1.020	0.19200000
Heredity	0.389	0.234	0.646	0.00026500

quences of interval cancer for women's health are more severe due to its higher biological aggressiveness.

The clinical significance of this model is that it provides a tool for categorizing women into higher or lower risk subgroups for developing interval breast cancer, which is a step toward more tailored screening [56].

Despite the fact that time is critical for outcomes, sending patients for additional testing is not always a well-organized process.

According to Bellio et al. (2017), the higher biological aggressiveness of interval cancer compared to screen-detected cancer requires the development of more sensitive imaging techniques as well as a specific diagnostic path for high-risk women [7].

Hofvind et al. (2018) suggest that women with false-positive mammography should be offered a new mammogram within 6 months or a year to detect cancer at an earlier stage, instead of after two years [57].

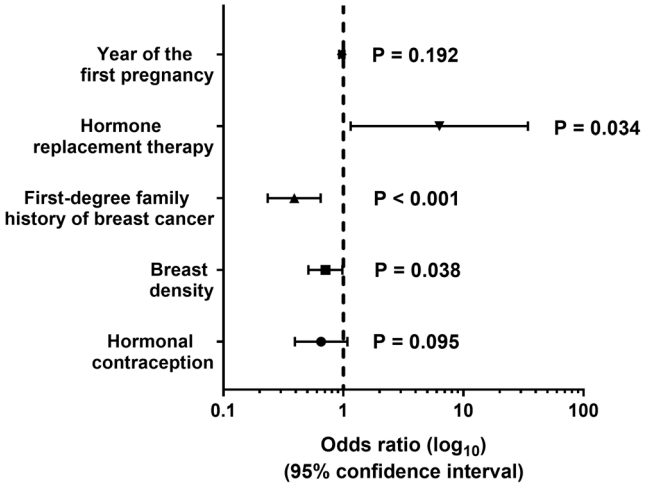

Figure 1. The odds ratio of the variables in the multivariate logistic regression model. The variables of breast tissue density and first-degree family history of breast cancer increase the likelihood of interval cancer, whereas hormone replacement therapy is positively related to screen-detected cancer.

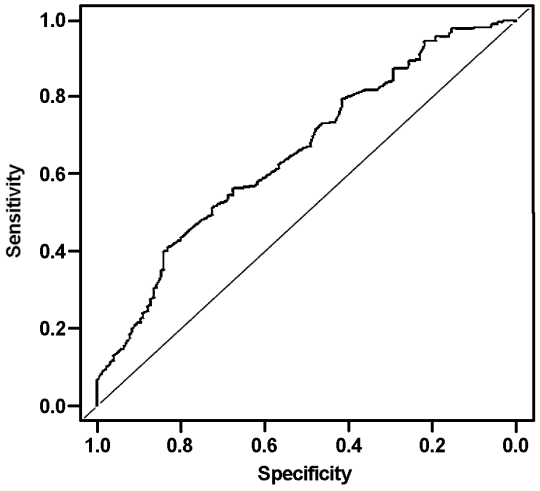

Figure 2. Area under the ROC curve 0.658 (95% confidence limit (CI) 0.602–0.713). The ROC curve describes the classification/prediction value of logistic regression. The area under the ROC curve is a measure of the accuracy of the test. If the logistic model is better than random sorting, then the ROC curve is shifted towards the upper left corner and has an area under the curve greater than 0.5. The prediction model has an area under the ROC curve of 0.658, with a 95% CI ranging from 0.602 to 0.713. It's considered that the test with CI between 0.60–0.70 is a weaker test.

Risk-based screening has the potential to reduce screening-related harms while increasing screening effectiveness [58].

Breast-imaging technologies such as ultrasonography, tomosynthesis, and MRI detect cancers in dense breasts that mammography misses [59].

Increased breast density can obscure cancers in dense tissue on mammography, resulting in lower mammographic accuracy. According to Lee et al. (2017), most women with dense breasts and no other risk factors would likely experience more harm than benefits from supplemental screening ultrasound; however, women with dense breasts and additional risk factors that place them at high lifetime risk of developing breast cancer (>20%) should undergo breast MRI rather than supplemental screening ultrasound [60, 61].

In high-risk women, MRI is more sensitive than screening mammography for detecting invasive breast cancers, possibly because these women are younger and have dense breasts [62].

Annual MRI combined with mammography screening for women at high risk for breast cancer has the potential to be effectively adopted into an organized breast screening program [63].

There are no specific guidelines for reporting on studies involving breast cancer risk prediction models [64, 65].

With a discriminant accuracy of 0.658 (95% CI 0.602–0.713), our interval cancer prediction model can be considered as a weaker, but acceptable predictive test. It has the advantage of being based on real-world data and rigorously adhering to the interval cancer criteria.

Although we are aware that existing breast cancer prediction models do not address interval cancer risk, the systematic review of prior breast cancer risk prediction models conducted by Anothaisintawee et al. is useful for gaining insight into the discriminative capacity of these models. The discriminatory power performance of most models was weak to be acceptable in both internal and external validation (concordance statistic: 0.53–0.66). The authors believe that the low discriminatory power of existing models is due to a lack of knowledge about risk factors, heterogeneous subtypes of breast cancer, and different risk factor distributions in different populations [65].

The concordance statistic, according to Caetano et al. (2018), only measures discrimination and not the calibration. As a result, when only a small number of patients had an event of interest, it is not a good measure of the true probability that a given patient experienced an event [66–69]. According to Mullooly et al. (2021), the occurrence of interval cancer must be validated with a time lag, which can often take many years [70]. The sample size, according to Altman and Royston, determines the model's characteristics. In studies with small sample sizes and a large number of risk factors in the model, there is a high likelihood that unimportant variables will be included while important variables will be excluded. Large-sample studies, on the other hand, are more likely to include statistically significant variables with no clinical significance. In our model, we included 789 participants with interval cancer and 3,530 with screen-detected cancer, and we believe that three screening cycles is a reasonable time frame for the occurrence of interval cancer and the use of concordance statistics.

Because the country's breast cancer screening program was just implemented and there were organizational issues associated with it, we observed that a lot of data were missing

in the screening database or were not submitted in a unique way. We believe that the lack of most family history data does not provide true insight into the First-degree family history of breast cancer variable as a risk factor for interval cancer.

Breast cancer risk prediction models, according to Lee et al. (2019), include different groups of risk factors that are weighted differently and may contribute to different outcomes for the same patient. Modeling can be useful in predicting events but the outcomes are heavily dependent on assumptions. The validity of the results is influenced by incorrect or uncertain assumptions. There is no model that is suitable for all subgroups of the general population [71]. To our consideration, our model is best suited for use in the Croatian population of women because we have identified the available risk factors for the development of interval cancer in our population, but with knowledge of the specific epidemiological picture, it can be applied more broadly. The conclusion stems from the fact that many of the newer models, particularly those based on established models, lack validation in cohorts other than their initial study populations (Chen model, Hispanic-Banegas, Tice, Barlow, Pankratz, Tworoger) and require additional validation before being used routinely in clinic and research settings [72].

In a model, we included highly selected variables that had previously demonstrated statistical significance at the 0.05 level in univariate logistic regression (breast density, year of the first pregnancy, hormonal contraception, and hormone replacement therapy).

The variables "breast density" and "hormone replacement therapy" were found to be statistically significant in both univariate and multivariate logistic regression. This is consistent with the literature on breast density [7, 24, 73–75] and on hormone replacement therapy [11, 14, 15, 76], and their mutual interaction [5, 75, 77–79].

We made an exception for a variable first-degree family history of breast cancer, despite the fact that the χ^2-test did not indicate a statistically significant difference between family history of breast cancer and cancer neither statistical significance was not confirmed by univariate logistic regression. We assume that it can be caused by a huge lack of data in the screening database related to the family history of breast cancer. Another reason for including this variable was that when we looked at consanguinity between participants, we discovered that first-degree family history of breast cancer was the most common type of family history in both interval and screen-detected cancer in our study (18% interval cancer vs. 82% screen-detected cancer). With such a noticeable prevalence, it is reasonable to expect that this predictor will follow the patterns of association with interval cancer that have been described in the literature [11, 73, 74, 80–83]. It is also well recognized that the number of affected relatives raises the risk of having interval cancer [11, 14, 15, 20, 73, 81, 82].

By analyzing screening data, we discovered that another risk factor, breast density, did not follow the patterns of association with interval cancer described in the literature. The literature describes a higher incidence and extremely aggressive phenotype of interval cancer in women with extremely high breast density ("50–75%" or ACR-D, extremely dense) [11, 73, 84] or with extremely low breast density ("25%" or ACR-A, almost entirely fatty) [12, 24, 85]. In our screening, the most common breast tissue density for both cancers were "25–50%" (ACR-B, scattered areas of fibroglandular density) (47.84%, interval cancer vs. 44.49%, screen-detected cancer), while high breast tissue density "50–75%" was the least represented (17.04%, interval cancer vs. 12.48% screen-detected cancer) despite the difference between breast density and cancer in our screening was statistically significant (χ^2=19.20, number of degrees of freedom = 2, p<0.001).

The variable "use of hormonal contraceptives" was interesting to us because of various literary discrepancies on the impact of hormonal contraception on breast cancer. On the one hand, it is claimed that hormonal contraception increases the risk of breast cancer [37–39], while on the other hand, this association is denied [33, 34]. Based on the literature [28–34, 36–39, 47], we can conclude that the effect of hormonal contraception on breast cancer risk is dependent on the variety of hormones in their composition. Although the χ^2-test revealed a statistically significant relationship between hormonal contraception use and both types of cancer (χ^2=4.21, number of degrees of freedom = 1, p=0.040), in the logistic model, a p-value of less than 0.1 (0.095) in the logistic model suggests that, in the context of recommendations, hormonal contraception may only have an effect on the increased risk of screen-detected cancer.

Data on certain additional major breast cancer risk factors, which may be incorporated in the model, cannot be collected from screening databases. According to Holm et al. and Bellio et al., interval cancer and screen-detected cancer have distinct genetic profiles [7, 12]. Grassmann et al. (2019) discovered that interval cancers are more likely than screen-detected cancers detected to have rare deleterious mutations in genes that may increase the risk of other non-breast cancer. The findings could have implications for a screening program [77].

A better understanding of risk factor mechanisms, as well as newly discovered risk factors, can be used to improve risk prediction models for interval breast cancer.

The decision to include a new risk factor in prediction models is based on relative costs as well as the potential for illness prevention and life-saving [86]. Lee et al. (2015) state that delaying or reducing the frequency of mammographic screening for women who do not qualify for restrictive risk-based screening approaches will almost certainly result in delayed cancer detection and sacrifice a significant portion of the screening's mortality-reducing, life-extending, and morbidity-reducing benefits [71].

The interval cancer risk model should be simple and easy to interpret in order to be used in routine screening. Neither the patients nor the family doctors have extensive mathe-

matical and statistical training to interpret complex models. Even though it will be useful, public health care providers in many countries lack the resources to include variables that require specialized genetic measurement techniques for wider screening [65].

In the case of a familial predisposition to breast cancer, prior genetic testing and monitoring (BRCA1 or BRCA2 mutation) would probably be more reasonable.

Environmental factors, lifestyle, health care availability and quality, and comorbidity could all be considered as model inputs.

According to the 2010 conclusion of the IARC Monography, the consumption of alcoholic beverages was causally associated with the occurrence of female breast cancer [87].

Physical activity can be included because of its relationship with BMI, which we used in the preceded univariate logistic regression [76, 88].

Han et al. (2017) conducted a meta-analysis that found a significant link between hypertension and breast cancer risk, particularly in postmenopausal hypertensive women. When the questionnaire asked about hypertension, it could be included in the model and will be useful because the mean age of women in our screening was 60.7±5.99 years, with a range of 54.71 to 66.69 years, implying that most women in menopause and postmenopause participated in our screening [89].

Connecting medical services in healthcare through digitalization could track the time between a false-positive mammogram finding and the next diagnostic procedure, increasing the accuracy of interval cancer detection by sending a signal to the screening program. These women could probably be better identified and monitored in screening.

A previous false-positive mammogram is also a known risk factor for interval cancer [11, 74, 77, 90]. The model could be populated with the previous false-positive mammograms. Women who have had a false-positive test have a higher risk of cancer detection in subsequent screenings, particularly those who have had a false-positive result involving cytology or biopsy [91].

According to Hofvind et al. (2012), interval cancer may be a recognizable but overlooked finding during screening in 3–35% of cases and the majority of interval cancers (2/3) were discovered in the second year of screening. Women who have a false-positive screening result may be offered a follow-up mammogram within 6 months or a year, rather than two years [92]. Even more important is to distinguish interval cancer from false-negative mammography findings [11]. According to Houssami and Hunter (2017), 20–25% of interval breast cancers are missed (false negative) [21]. Fong et al. (2014) demonstrated that the mortality prognosis of true interval cancers is similar to that of screen-detected cancers, whereas it is much worse for false-negative cancers [93]. We agree with Blanch (2014), who considers early rescreening as a protective factor in interval cancer [11]. In our study, we did not observe the type of interval cancer at screening or the association of interval cancer with other tumors. Therefore, our work leaves the possibility of further improving the model with these data.

Future risk models, according to Olsson and Olsson (2020), should take into account the length of the menstrual cycle, its regularity, the number of cycles prior to the first pregnancy, and the number of cycles throughout life [46]. Related to the breast density in our model, we consider that these factors are useful in predicting interval cancer. Pike et al. modeled age-specific breast cancer incidence as a function of breast tissue aging, which summed up the effects of estrogen and progesterone exposure [94, 95].

Different measurements of mammographic density, which has shown a favorable correlation with interval cancer, can be used to improve the model. Instead of using the BI-RADS tool, some of the two innovative mammogram-based breast cancer risk factors based on image brightness (Cirrocumulus) and texture (Cirrus) can be employed to make this variable, as well as a whole model, more robust and objective. New measures, according to Nguyen et al., appear to be more strongly related to breast cancer causative factors than conventional mammographic density, implying that a woman's mammography contains more risk information than her genome. Cirrocumulus does better at identifying women at higher-than-average risk, while Cirrus does better at identifying women at lower-than-average risk. Finding new approaches to extract risk information from mammograms could pave the way for risk-based, individualized breast screening [96]. Another option is to employ artificial intelligence (AI) system instead of BI-RADS in the future [97].

Having a first-degree relative with breast cancer was linked to interval cancer in our model, and using hormone replacement therapy was linked to an increased risk of screen-detected cancer. With a high level of uncertainty, the model could be used as an additional tool in risk assessment in women with Hereditary Breast and Ovarian Cancer Syndrome.

Holm et al. (2015) believe that there is different genetic background of interval cancer and screening-detected cancer. Despite this fact, the Myriad II model and The Tyrer-Cuzick (TC) model (AKA: International Breast Cancer Study (IBIS) Breast Cancer Risk Evaluation Tool) include hereditary data, such as first- and second-degree relatives with breast and/or ovarian cancer [72].

Furthermore, Grassman et al. discovered that the association of family history of other cancers with interval cancer risk suggests that rare penetrant cancer mutations predispose individuals to both interval breast cancer and other cancers and that the additional large-scale sequencing efforts (particularly in patients with multiple tumors) are required to uncover the underlying cause of the observed associations [77].

Although our model has lower predictor reliability, it can be used to make recommendations for individual screening participants. A first-degree family history of breast cancer and

breast density indicate an increased risk of detecting interval cancer between mammograms, and individualized risk screening should be considered (modification of screening interval or additional screening by magnetic resonance or ultrasound). Furthermore, participants who use hormone replacement therapy must be medically monitored due to the increased risk of screen-detected cancer during screening. Participants in the screening program who use hormone replacement therapy and have a higher density of breast tissue should be encouraged to have more frequent mammograms.

Study limitations. Historically, retrospective studies have had a significant impact on clinical practice when it comes to associations between risk factors and outcomes (e.g., the association between smoking and lung cancer) [98]. As per Talari and Goyal (2020), they are useful in studying rare diseases and rare outcomes [99].

Our study is a retrospective analysis of the screening cohort and its design implies that there is no selection bias. Despite the fact that women with a family history of breast cancer are more likely to get screened, this could be a confounding factor when results are extrapolated to the entire female population.

Furthermore, the variables in the screening database were predetermined, limiting the number of variables that could be used to build the model.

The incompleteness of some data was obvious because a breast cancer screening program was about starting with the associated beginning data processing challenges. Methods such as data mining and imputation can be used. BI-RADS as a tool involves radiologist subjectivity in judgments and the chance of false-negative mammograms, even double reading. Interval cancers were not classified in our study. Our model's predictor reliability hasn't been tested in another similar cohort.

References

[1] EUROPEAN COMMISSION. Directorate-General for Health and Consumer Protection NP. European guidelines for quality assurance in breast cancer screening and diagnosis [Internet]. 4th ed. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R von KL, Technical editors. Luxembourg: Office for Official Publications of the European Communities; 2006. pp. 416. https://www.euref.org/european-guidelines/4th-edition

[2] AUTIER P, BONIOL M, LAVECCHIA C, VATTEN L, GAVIN A et al. Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database. BMJ 2010; 341:c 3620. Erratum in: BMJ 2010; 341: c4480. LaVecchia, Carlo [corrected to La Vecchia, Carlo]. https://doi.org/10.1136/bmj.c3620

[3] TABÁR L, VITAK B, CHEN TH, YEN AM, COHEN A et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology 2011; 260: 658–663. https://doi.org/10.1148/radiol.11110469

[4] ZIELONKE N, GINI A, JANSEN EEL, ANTTILA A, SEGNAN N et al. Evidence for reducing cancer-specific mortality due to screening for breast cancer in Europe: A systematic review. Eur J Cancer 2020; 127: 191–206. https://doi.org/10.1016/j.ejca.2019.12.010

[5] AUTIER P, BONIOL M. Mammography screening: A major issue in medicine. Eur J Cancer 2018; 90: 34–62. https://doi.org/10.1016/j.ejca.2017.11.002

[6] HEIDINGER O, BATZLER WU, KRIEG V, WEIGEL S, BIESHEUVEL C et al. The incidence of interval cancers in the German mammography screening program: results from the population-based cancer registry in North Rhine-Westphalia. Dtsch Arztebl Int 2012; 109: 781–787. https://doi.org/10.3238/arztebl.2012.0781

[7] BELLIO G, MARION R, GIUDICI F, KUS S, TONUTTI M et al. Interval Breast Cancer Versus Screen-Detected Cancer: Comparison of Clinicopathologic Characteristics in a Single-Center Analysis. Clin Breast Cancer. 2017; 17: 564–571. https://doi.org/10.1016/j.clbc.2017.04.001

[8] MESHKAT B, PRICHARD RS, AL-HILLI Z, BASS GA, QUINN C et al. A comparison of clinical-pathological characteristics between symptomatic and interval breast cancer. Breast 2015; 24: 278–282. https://doi.org/10.1016/j.breast.2015.02.032

[9] DEFOSSEZ G, QUILLET A, INGRAND P. Aggressive primary treatments with favourable 5-year survival for screen-interval breast cancers. BMC Cancer 2018; 18: 393. https://doi.org/10.1186/s12885-018-4319-4

[10] DUFFY SW, DIBDEN A, MICHALOPOULOS D, OFFMAN J, PARMAR Det al. Screen detection of ductal carcinoma in situ and subsequent incidence of invasive interval breast cancers: a retrospective population-based study. Lancet Oncol 2016; 17: 109–14. Erratum in: Lancet Oncol 2016; 17: e46. https://doi.org/10.1016/S1470-2045(15)00446-5

[11] BLANCH J, SALA M, IBÁÑEZ J, DOMINGO L, FERNANDEZ B et al. Impact of risk factors on different interval cancer subtypes in a population-based breast cancer screening programme. PLoS One 2014; 9: e110207. https://doi.org/10.1371/journal.pone.0110207

[12] HOLM J, HUMPHREYS K, LI J, PLONER A, CHEDDAD A et al. Risk factors and tumor characteristics of interval cancers by mammographic density. J Clin Oncol 2015; 33: 1030–1037. https://doi.org/10.1200/JCO.2014.58.9986

[13] FAJ D, POSEDEL D, STIMAC D, IVEZIC Z, KASABASIC M et al. Survey of mammography practice in Croatia: equipment performance, image quality and dose. Radiat Prot Dosimetry 2008; 131: 535–540. https://doi.org/10.1093/rpd/ncn283

[14] DOMINGO L, BLANCH J, SERVITJA S, COROMINAS JM, JOSEP M et al. Aggressiveness features and outcomes of true interval cancers: comparison between screen-detected and symptom-detected cancers. Eur J Cancer Prev 2013; 22: 21–28. https://doi.org/10.1097/CEJ.0b013e328354d324

[15] KIRSH VA, CHIARELLI AM, EDWARDS SA, O'MALLEY FP, SHUMAK RS et al. Tumor characteristics associated with mammographic detection of breast cancer in the Ontario breast screening program. J Natl Cancer Inst 2011; 103: 942–950. https://doi.org/10.1093/jnci/djr138

[16] COLLABORATIVE GROUP ON HORMONAL FACTORS IN BREAST CANCER. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet 2001; 358: 1389–1399. https://doi.org/10.1016/S0140-6736(01)06524-2

[17] KERLIKOWSKE K, ZHU W, TOSTESON AN, SPRAGUE BL, TICE JA et al. Identifying women with dense breasts at high risk for interval cancer: a cohort study. Ann Intern Med 2015; 162: 673–681. https://doi.org/10.7326/M14-1465

[18] KHARAZMI E, CHEN T, NAROD S, SUNDQUIST K, HEMMINKI K. Effect of multiplicity, laterality, and age at onset of breast cancer on familial risk of breast cancer: a nationwide prospective cohort study. Breast Cancer Res Treat 2014; 144: 185–192. https://doi.org/10.1007/s10549-014-2848-3

[19] NELSON HD, ZAKHER B, CANTOR A, FU R, GRIFFIN J et al. Risk factors for breast cancer for women aged 40 to 49 years: a systematic review and meta-analysis. Ann Intern Med 2012; 156: 635–648. https://doi.org/10.7326/0003-4819-156-9-201205010-00006

[20] ROMÁN M, QUINTANA MJ, FERRER J, SALA M, CASTELLS X. Cumulative risk of breast cancer screening outcomes according to the presence of previous benign breast disease and family history of breast cancer: supporting personalised screening. Br J Cancer 2017; 116: 1480–1485. https://doi.org/10.1038/bjc.2017.107

[21] HOUSSAMI N, HUNTER K. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. NPJ Breast Cancer 2017; 3: 12. https://doi.org/10.1038/s41523-017-0014-x

[22] EVANS DG, HOWELL A. Can the breast screening appointment be used to provide risk assessment and prevention advice? Breast Cancer Res 2015; 17: 84. https://doi.org/10.1186/s13058-015-0595-y

[23] STRAND F, HUMPHREYS K, HOLM J, ERIKSSON M, TÖRNBERG S et al. Long-term prognostic implications of risk factors associated with tumor size: a case study of women regularly attending screening. Breast Cancer Res 2018; 20: 31. https://doi.org/10.1186/s13058-018-0962-6

[24] BOYD NF, HUSZTI E, MELNICHOUK O, MARTIN LJ, HISLOP G et al. Mammographic features associated with interval breast cancers in screening programs. Breast Cancer Res 2014; 16: 417. https://doi.org/10.1186/s13058-014-0417-7

[25] HSIEH CC, TRICHOPOULOS D, KATSOUYANNI K, YUASA S. Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: associations and interactions in an international case-control study. Int J Cancer 1990; 46: 796–800. https://doi.org/10.1002/ijc.2910460508

[26] KELSEY JL, GAMMON MD, JOHN EM. Reproductive factors and breast cancer. Epidemiol Rev 1993; 15: 36–47.

[27] COLLABORATIVE GROUP ON HORMONAL FACTORS IN BREAST CANCER, GOLDBOHM RA, VAN DEN BRANDT P. Type and timing of menopausal hormone therapy and breast cancer risk: individual participant meta-analysis of the worldwide epidemiological evidence. Lancet 2019; 394: 1159–1168. https://doi.org/10.1016/S0140-6736(19)31709-X

[28] MACMAHON B, COLE P, LIN TM, LOWE CR, MIRRA AP et al. Age at first birth and breast cancer risk. Bull World Health Organ 1970; 43: 209–221.

[29] ALBREKTSEN G, HEUCH I, HANSEN S, KVÅLE G. Breast cancer risk by age at birth, time since birth and time intervals between births: Exploring interaction effects. Br J Cancer 2005; 92: 167–175. https://doi.org/10.1038/sj.bjc.6602302

[30] MERRILL RM, FUGAL S, NOVILLA LB, RAPHAEL MC. Cancer risk associated with early and late maternal age at first birth. Gynecol Oncol 2005; 963: 583–593. https://doi.org/10.1016/j.ygyno.2004.11.038

[31] CALLIHAN EB, GAO D, JINDAL S, LYONS TR, MANTHEY E et al. Postpartum diagnosis demonstrates a high risk for metastasis and merits an expanded definition of pregnancy-associated breast cancer. Breast Cancer Res Treat 2013; 138: 549–559. https://doi.org/10.1007/s10549-013-2437-x

[32] SLEPICKA PF, CYRILL SL, DOS SANTOS CO. Pregnancy and Breast Cancer: Pathways to Understand Risk and Prevention. Trends Mol Med 2019; 25: 866–881. https://doi.org/10.1016/j.molmed.2019.06.003

[33] GAFFIELD ME, CULWELL KR, RAVI A. Oral contraceptives and family history of breast cancer. Contraception 2009; 80: 372–380. https://doi.org/10.1016/j.contraception.2009.04.010

[34] WORLD HEALTH ORGANIZATION. Medical eligibility criteria for contraceptive use. Fifth edit. Geneva; 2015. ISBN-13: 978-92-4-154915-8 https://www.ncbi.nlm.nih.gov/books/NBK321151/

[35] SUN X, NICHOLS HB, TSE CK, BELL MB, ROBINSON WR et al. Association of Parity and Time since Last Birth with Breast Cancer Prognosis by Intrinsic Subtype. Cancer Epidemiol Biomarkers Prev 2016; 25: 60–7. https://doi.org/10.1158/1055-9965.EPI-15-0864

[36] HUSBY A, WOHLFAHRT J, ØYEN N, MELBYE M. Pregnancy duration and breast cancer risk. Nat Commun 2018; 9: 4255. https://doi.org/10.1038/s41467-018-06748-3

[37] MØRCH LS, SKOVLUND CW, HANNAFORD PC, IVERSEN L, FIELDING S et al. Contemporary Hormonal Contraception and the Risk of Breast Cancer. N Engl J Med 2017; 377: 2228–2239. https://doi.org/10.1056/NEJMoa1700732

[38] BEABER EF, BUIST DS, BARLOW WE, MALONE KE, REED SD et al. Recent oral contraceptive use by formulation and breast cancer risk among women 20 to 49 years of age. Cancer Res 2014; 74: 4078–4089. https://doi.org/10.1158/0008-5472.CAN-13-3400

[39] WHITE ND. Hormonal Contraception and Breast Cancer Risk. Am J Lifestyle Med 2018; 12: 224–226. https://doi.org/10.1177/1559827618754833

[40] NGUYEN TL, AUNG YK, LI S, TRINH NH, EVANS CF et al. Predicting interval and screen-detected breast cancers from mammographic density defined by different brightness thresholds. Breast Cancer Res 2018; 20: 1–11. https://doi.org/10.1186/s13058-018-1081-0

[41] GOTTSCHALK MS, ESKILD A, HOFVIND S, GRAN JM, BJELLAND EK. Temporal trends in age at menarche and age at menopause: a population study of 312656 women in Norway. Hum Reprod 2020; 35: 464–471. https://doi.org/10.1093/humrep/dez288

[42] VINOGRADOVA Y, COUPLAND C, HIPPISLEY-COX J. Use of hormone replacement therapy and risk of breast cancer: nested case-control studies using the QResearch and CPRD databases. BMJ 2020; 371: m3873. https://doi.org/10.1136/bmj.m3873
[43] ALEXANDER FE, ROBERTS MM. The menopause and breast cancer. J Epidemiol Community Health 1987; 41: 94–100. https://doi.org/10.1136/jech.41.2.94
[44] 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 2012; 13: 1141–1151. https://doi.org/10.1016/S1470-2045(12)70425-4
[45] GOLDBERG M, D'ALOISIO AA, O'BRIEN KM, ZHAO S, SANDLER DP. Pubertal timing and breast cancer risk in the Sister Study cohort. Breast Cancer Res 2020; 22: 112. https://doi.org/10.1186/s13058-020-01326-2
[46] OLSSON HL, OLSSON ML. The Menstrual Cycle and Risk of Breast Cancer: A Review. Front Oncol 2020; 10: 21. https://doi.org/10.3389/fonc.2020.00021
[47] CHARLTON BM, RICH-EDWARDS JW, COLDITZ GA, MISSMER SA, ROSNER BA et al. Oral contraceptive use and mortality after 36 years of follow-up in the Nurses' Health Study: prospective cohort study. BMJ 2014; 349: g6356. https://doi.org/10.1136/bmj.g6356
[48] GILMOUR SG. The Interpretation of Mallows's Cp-Statistic. The Statistician 1996; 45: 49–56. https://doi.org/10.2307/2348411
[49] STEEL SJ, UYS DW. Variable selection in multiple linear regression: The influence of individual cases. ORiON 2007; 23: 123–136. https://doi.org/10.5784/23-2-52
[50] KIM C, HWANG S. Influential subsets on the variable selection. Commun Stat Theory Methods 2000; 29: 335–347. https://doi.org/10.1080/03610920008832487
[51] ZHANG Z. Variable selection with stepwise and best subset approaches. Ann Transl Med 2016; 4: 136. https://doi.org/10.21037/atm.2016.03.35
[52] JUNHUI L, XIAOHUAN L, KUN C WL. Stepwise Regression Analysis. Package 'StepReg' 2022. https://cran.r-project.org/web/packages/StepReg/StepReg.pdf
[53] RODRIGUEZ G. Introducing R. Online Version. 5. Generalized Linear Models. https://data.princeton.edu/r/glms
[54] NGUYEN TL, LI S, DITE GS, AUNG YK, EVANS CF, TRINH HN, et al. Interval breast cancer risk associations with breast density, family history and breast tissue aging. Int J Cancer 2020; 147: 375–382. https://doi.org/10.1002/ijc.32731
[55] ROSNER B, COLDITZ GA, WILLETT WC. Reproductive risk factors in a prospective study of breast cancer: The nurses' health study. Am J Epidemiol 1994; 139: 819–835. https://doi.org/10.1093/oxfordjournals.aje.a117079
[56] TRENTHAM-DIETZ A, KERLIKOWSKE K, STOUT NK, MIGLIORETTI DL, SCHECHTER CB et al. Tailoring breast cancer screening intervals by breast density and risk for women aged 50 years or older: Collaborative modeling of screening outcomes. Ann Intern Med 2016; 165: 700–712. https://doi.org/10.7326/M16-0476
[57] HOFVIND S, SAGSTAD S, SEBUØDEGÅRD S, CHEN Y, ROMAN M, LEE CI et al. Interval breast cancer rates and histopathologic tumor characteristics after false-positive findings at mammography in a populationbased screening program. Radiology 2018; 287: 58–67. https://doi.org/10.1148/radiol.2017162159
[58] BRAWLEY O. Editorial Annals of Internal Medicine Risk-Based Mammography Screening : An Effort to Maximize the Benefits and Minimize the Harms. Ann Intern Med 2012; 156: 662–663. https://doi.org/10.7326/0003-4819-156-9-201205010-00012
[59] TAGLIAFICO AS, CALABRESE M, MARISCOTTI G, DURANDO M, TOSTO S et al. Adjunct screening with tomosynthesis or ultrasound in women with mammography-negative dense breasts: Interim report of a prospective comparative trial. J Clin Oncol 2016; 34: 1882–1888. https://doi.org/10.1200/JCO.2015.63.4147
[60] LEE CI, CHEN LE, ELMORE JG. Risk-based Breast Cancer Screening: Implications of Breast Density. Med Clin North Am 2017; 101: 725–741. https://doi.org/10.1016/j.mcna.2017.03.005
[61] POSSO M, LOURO J, SÁNCHEZ M, ROMÁN M, VIDAL C, SALA M et al. Mammographic breast density: How it affects performance indicators in screening programmes? Eur J Radiol 2019; 110: 81–87. https://doi.org/10.1016/j.ejrad.2018.11.012
[62] KRIEGE M, BREKELMANS CTM, BOETES C, BESNARD PE, ZONDERLAND HM et al. Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. N Engl J Med 2004; 351: 427–437. https://doi.org/10.1056/NEJMoa031759
[63] CHIARELLI AM, PRUMMEL M V., MURADALI D, MAJPRUZ V, HORGAN M et al. Effectiveness of screening with annual magnetic resonance imaging and mammography: Results of the initial screen from the Ontario High Risk Breast Screening Program. J Clin Oncol 2014; 32: 2224–2230. https://doi.org/10.1200/JCO.2013.52.8331
[64] STEYERBERG EW, VICKERS AJ, COOK NR, GERDS T, GONEN M et al. Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology 2010; 21: 128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2
[65] ANOTHAISINTAWEE T, TEERAWATTANANON Y, WIRATKAPUN C, KASAMESUP V, THAKKINSTIAN A. Risk prediction models of breast cancer: A systematic review of model performances. Breast Cancer Res Treat 2012; 133: 1–10. https://doi.org/10.1007/s10549-011-1853-z
[66] CAETANO SJ, SONPAVDE G, POND GR. C-statistic: A brief explanation of its construction, interpretation and limitations. Eur J Cancer 2018; 90: 130–132. https://doi.org/10.1016/j.ejca.2017.10.027
[67] AUSTIN PC, STEYERBERG EW. Interpreting the concordance statistic of a logistic regression model: Relation to the variance and odds ratio of a continuous explanatory variable. BMC Med Res Methodol 2012; 12: 82. https://doi.org/10.1186/1471-2288-12-82

[68] SNELL KIE, ARCHER L, ENSOR J, BONNETT LJ, DEBRAY TPA et al. External validation of clinical prediction models: simulation-based sample size calculations were more reliable than rules-of-thumb. J Clin Epidemiol 2021; 135: 79–89. https://doi.org/10.1016/j.jclinepi.2021.02.011

[69] AUSTIN PC, STEYERBERG EW. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. Stat Med 2019; 38: 4051–4065. https://doi.org/10.1002/sim.8281

[70] MULLOOLY M, WHITE G, BENNETT K, O'DOHERTY A, FLANAGAN F et al. Retrospective radiological review and classification of interval breast cancers within population-based breast screening programmes for the purposes of open disclosure: A systematic review. Eur J Radiol 2021; 138: 109572. https://doi.org/10.1016/j.ejrad.2021.109572

[71] LEE CS, SICKLES EA, MOY L. Risk stratification for screening mammography: Benefits and harms. AJR Am J Roentgenol 2019; 212: 250–258. https://doi.org/10.2214/AJR.18.20345

[72] CINTOLO-GONZALEZ JA, BRAUN D, BLACKFORD AL, MAZZOLA E, ACAR A, et al. Breast cancer risk models: a comprehensive overview of existing models, validation, and clinical applications. Breast Cancer Res Treat 2017; 164: 263–284.

[73] HOUSSAMI N, ABRAHAM LA, KERLIKOWSKE K, BUIST DSM, IRWIG L et al. Risk factors for second screen-detected or interval breast cancers in women with a personal history of breast cancer participating in mammography screening. Cancer Epidemiol Biomarkers Prev 2013; 22: 946–961. https://doi.org/10.1158/1055-9965.EPI-12-1208-T

[74] ROTHSCHILD J, LOURENCO AP, MAINIERO MB. Screening mammography recall rate: Does practice site matter? Radiology 2013; 269: 348–353. https://doi.org/10.1148/radiol.13121487

[75] AZAM S, LANGE T, HUYNH S, ARO AR, VON EULER-CHELPIN M et al. Hormone replacement therapy, mammographic density, and breast cancer risk: a cohort study. Cancer Causes Control 2018; 29: 495–505. https://doi.org/10.1007/s10552-018-1033-0

[76] PIZOT C, BONIOL M, MULLIE P, KOECHLIN A, BONIOL M et al. Physical activity, hormone replacement therapy and breast cancer risk: A meta-analysis of prospective studies. Eur J Cancer 2016; 52: 138–154. https://doi.org/10.1016/j.ejca.2015.10.063

[77] GRASSMANN F, HE W, ERIKSSON M, GABRIELSON M, HALL P et al. Interval breast cancer is associated with other types of tumors. Nat Commun 2019; 10: 4648. https://doi.org/10.1038/s41467-019-12652-1

[78] YOO TK, HAN KD, KIM D, AHN J, PARK WC et al. Hormone replacement therapy, breast cancer risk factors, and breast cancer risk: a nationwide population-based cohort. Cancer Epidemiol Biomarkers Prev 2020; 297: 1341–1347. https://doi.org/10.1158/1055-9965.EPI-20-0038

[79] STONE J, DITE GS, GUNASEKARA A, ENGLISH DR, MCCREDIE MR et al. The heritability of mammographically dense and nondense breast tissue. Cancer Epidemiol Biomarkers Prev 2006; 15: 612–617. https://doi.org/10.1158/1055-9965.EPI-05-0127

[80] MAJEK O, DANES J, SKOVAJSOVA M, BARTONKOVA H, BURESOVA L et al. Breast cancer screening in the Czech Republic: time trends in performance indicators during the first seven years of the organised programme. BMC Public Health 2011; 11: 288. https://doi.org/10.1186/1471-2458-11-288

[81] SHIYANBOLA OO, ARAO RF, MIGLIORETTI DL, SPRAGUE BL, HAMPTON JM et al. Emerging trends in family history of breast cancer and associated risk. Cancer Epidemiol Biomarkers Prev 2017; 26: 1753–1760. https://doi.org/10.1158/1055-9965.EPI-17-0531

[82] MOHD RAZIF SM, SULAIMAN S, HANIE SS, AINA EN, ROHAIZAK M et al. The contribution of reproductive factors and family history towards premenopausal breast cancer risk in Kuala Lumpur, Malaysia. Med J Malaysia 2011; 66: 220–226.

[83] BARLOW WE, WHITE E, BALLARD-BARBASH R, VACEK PM, TITUS-ERNSTOFF L et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. J Natl Cancer Inst 2006; 98: 1204–1214. https://doi.org/10.1093/jnci/djj331

[84] STONE J. Should breast cancer screening programs routinely measure mammographic density? J Med Imaging Radiat Oncol 2018; 62: 151–158. https://doi.org/10.1111/1754-9485.12652

[85] SALA M, DOMINGO L, LOURO J, TORÁ-ROCAMORA I, BARÉ M et al. Survival and disease-free survival by breast density and phenotype in interval breast cancers. Cancer Epidemiol Biomarkers Prev 2018; 27: 908–916. https://doi.org/10.1158/1055-9965.EPI-17-0995

[86] GAIL MH, PFEIFFER RM. On criteria for evaluating models of absolute risk. Biostatistics 2005; 6: 227–239. https://doi.org/10.1093/biostatistics/kxi005

[87] WORLD HEALTH ORGANIZATION INTERNATIONAL AGENCY FOR RESEARCH ON CANCER Carbon black, titanium dioxide, and talc / IARC Working Group on the Evaluation of Carcinogenic Risks to Humans: IARC monographs on the evaluation of carcinogenic risks to humans; vol. 93. IARC Lyon 2006, France ISBN 978 92 832 1293 5

[88] 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. https://doi.org/10.1007/s10549-011-1818-2

[89] HAN H, GUO W, SHI W, YU Y, ZHANG Y et al. Hypertension and breast cancer risk: A systematic review and meta-analysis. Sci Rep 2017; 7: 1–9. https://doi.org/10.1038/srep44877

[90] DIMITROVA N, ZNAOR A, AGIUS D, ESER S, SEKERIJA M et al. Breast cancer in South-Eastern European countries since 2000: Rising incidence and decreasing mortality at young and middle ages. Eur J Cancer 2017; 83: 43–55. https://doi.org/10.1016/j.ejca.2017.06.011

[91] CASTELLS X, ROMÁN M, ROMERO A, BLANCH J, ZUBIZARRETA R et al. Breast cancer detection risk in screening mammography after a false-positive result. Cancer Epidemiol 2013; 37: 85–90. https://doi.org/10.1016/j.canep.2012.10.004

[92] HOFVIND S, SKAANE P. Stage distribution of breast cancer diagnosed before and after implementation of population-based mammographic screening. Rofo 2012; 184: 437–442. https://doi.org/10.1055/s-0031-1299352

[93] FONG Y, EVANS J, BROOK D, KENKRE J, JARVIS P et al. The incidence and 10-year survival of interval breast cancers in Wales. Clin Radiol 2014; 69: e168–172. https://doi.org/10.1016/j.crad.2013.11.018

[94] PIKE MC, KRAILO MD, HENDERSON BE, CASAGRANDE JT, HOEL DG. "Hormonal" risk factors, "Breast tissue age" and the age-incidence of breast cancer. Nature 1983; 303: 767–770. https://doi.org/10.1038/303767a0

[95] RIAZ N, WOLDEN SL, GELBLUM DY, ERIC J, LOWERY JT et al. Differential expression of prognostic biomarkers between interval and screen-detected breast cancers: does age or family history matter? Breast Cancer Res Treat 2011; 129: 2011–2019. https://doi.org/10.1007/s10549-011-1448-8

[96] NGUYEN TL, SCHMIDT DF, MAKALIC E, MASKARINEC G, LI S et al. Novel mammogram-based measures improve breast cancer risk prediction beyond an established mammographic density measure. Int J Cancer 2021; 148: 2193–2202. https://doi.org/10.1002/ijc.33396

[97] MCKINNEY SM, SIENIEK M, GODBOLE V, GODWIN J, ANTROPOVA N et al. International evaluation of an AI system for breast cancer screening. Nature 2020; 577: 89–94. https://doi.org/10.1038/s41586-019-1799-6

[98] DOLL R, HILL AB. Smoking and carcinoma of the lung: Preliminary report. Bull World Health Organ 1999; 77: 84–93. https://doi.org/10.1136/bmj.2.4682.739

[99] TALARI K, GOYAL M. Retrospective studies – Utility and caveats. J R Coll Physicians Edinb 2020; 50: 398–402. https://doi.org/10.4997/JRCPE.2020.409