Cancer Epidemiology 86 (2023) 102448

Contents lists available at ScienceDirect

Cancer Epidemiology

journal homepage: www.elsevier.com/locate/canep

A brain metastasis prediction model in women with breast cancer

Bernardo Cacho-Díaz [a,1], Antelmo A. Meneses-García [b], Sergio I. Valdés-Ferrer [c,e,2], Nancy Reynoso-Noverón [d,*]

[a] Programa de Maestría y Doctorado en Ciencias Médicas, Odontológicas y de la Salud. Universidad Autónoma de Mexico (UNAM), ZC 04510, Mexico
[b] General Director. Instituto Nacional de Cancerología, Mexico City ZC 14080, Mexico
[c] Institute of Bioelectronic Medicine, Feinstein Institutes for Medical Research, Manhasset, NY, USA
[d] Research Unit. Instituto Nacional de Cancerología, Mexico City ZC 14080, Mexico
[e] Departamento de Neurologíay Psiquiatría, Instituto Nacional de Ciencias Médicas y Nutrición SalvadorZubiran, Mexico City ZC 14080, Mexico

ARTICLE INFO

Keywords:
Breast cancer
Brain metastases
Prediction model
Estrogen receptor status
Ki-67

ABSTRACT

Background: Breast cancer (BC) is a leading cause of mortality and the most frequent malignancy in women, and most deaths are due to metastatic disease, particularly brain metastases (BM). Currently, no biomarker or prediction model is used to predict BM accurately. The objective was to generate a BM prediction model from variables obtained at BC diagnosis.

Methods: A retrospective cohort of women with BC diagnosed from 2009 to 2020 at a single center was divided into a training dataset (TD) and a validation dataset (VD). The prediction model was generated in the TD, and its performance was measured in the VD using the area under the curve (AUC) and C-statistic.

Results: The cohort (n = 5009) was divided into a TD (n = 3339) and a VD (n = 1670). In the TD, the model with the best performance (lowest AIC) was built with the following variables: age, estrogen receptor status, tumor size, axillary adenopathy, anatomic clinical stage, Ki-67 expression, and Scarff–Bloom–Richardson score. This model had an AUC of 0.79 (95%CI, 0.76–0.82; p < 0.0001) in the TD. The 10-fold cross-validation showed the good stability of the model. The model displayed an AUC of 0.81 (95%CI, 0.77–0.85; P < 0.0001) in the VD. Four groups, according to the risk of BM, were generated. In the low-risk group, 1.2% were diagnosed with BM (reference); in the medium-risk group, 5.0% [HR 4.01 (95%CI, 1.8 – 8.8); P < 0.0001]; in the high-risk group, 8.5% [HR 8.33 (95%CI, 4.1–17.1); P < 0.0001]; and in the very high-risk group, 23.7% [HR 29.72 (95%CI, 14.9 – 59.1); P < 0.0001].

Conclusion: This prediction model built with clinical and pathological variables at BC diagnosis demonstrated robust performance in determining the individual risk of BM among patients with BC, but external validation in different cohorts is needed.

Importance of the study

A multivariable brain metastasis prediction model was constructed in a cohort of 5009 patients with breast cancer. The model demonstrated a remarkable discrimination capacity with an AUC of 0.81 (95%CI, 0.77–0.85; P < 0.0001) in the validation dataset.

1. Introduction

In women, breast cancer (BC) is the most common malignant tumor and the second cause of cancer-related deaths [1]. The reported leading cause of death in BC is metastatic disease [2], and of all metastatic disease locations, brain metastases (BM) are associated with the worst survival [3,4]. The occurrence of BM varies from 30% to 65% [5–7] in

Abbreviations: ACS, Anatomic clinical stage; AIC, Akaike information criteria; AUC, Area under the curve; BC, Breast cancer; BM, Brain metastases; ER, Estrogen receptor; HER2, Human Epidermal Growth Factor Receptor type 2; ROC, Receiver operating characteristic curve; SBR, Scarff–Bloom–Richardson score; TD, Training dataset; VD, Validation dataset.

* Correspondence to: Researcher in medical sciences. Research Unit. Instituto Nacional de Cancerología, Av. San Fernando 22. Col. Sección XVI, Tlalpan, Mexico City, Mexico.

E-mail address: nrn231002@yahoo.com.mx (N. Reynoso-Noverón).

[1] ORCID ID: 0000-0001-9289-5312

[2] ORCID ID: 0000-0002-4863-6484.

https://doi.org/10.1016/j.canep.2023.102448

Received 5 May 2023; Received in revised form 4 August 2023; Accepted 23 August 2023

postmortem studies to 0.2–1% in sizable epidemiologic studies [8].

Described risk factors for brain metastases in BC include age, tumor size, nodal involvement (especially axillary adenopathy), anatomic clinical stage, estrogen receptor (ER) expression status, progesterone receptor (PR) expression status, human epidermal growth factor receptor type 2 (HER2, c-erbB-2) overexpression, histologic grade as determined by the Scarff–Bloom–Richardson (SBR) grading score, and time to first metastasis [9,10]. Previous attempts to generate a BM prediction model include a nomogram to predict BM in metastatic breast cancer. This nomogram – built with the variables age, SBR, Status for ER, PR, and HER2, the delay between diagnosis and first metastasis, and the number of brain metastatic sites – presented an area under the receiver operating characteristic curve (AUC) of 0.68 in their training dataset (TD) and an AUC of 0.74 in their validation dataset (VD) [10]. Its external validation reported an AUC of 0.69 [11]. Currently, there is no individual biomarker to predict BM or a prediction model for all women with BC (not only those with metastatic disease) with good performance (AUC ideally $\geq$ 0.70). Therefore, we designed a study to build a BM prediction model from markers obtained at BC diagnosis to help in determining the odds of BM.

2. Methods

2.1. Data collection

All consecutive patients diagnosed with breast cancer between May 2009 and May 2020 were identified from the clinical archive department at the National Cancer Institute, Mexico. Trained registrars were prepared for data acquisition and were supervised weekly to review patients' electronic clinical charts.

2.2. Study population

To be included in the study, participants had to be female and over 18 years old, with an anatomic clinical stage (Stage) described at the time of initial diagnosis and confirmation of their ER, PR, and HER2 status by the institution's pathology department. Exclusion criteria comprised a previous diagnosis of neurological disease (such as epilepsy, congenital neural defects, neuroinfectious disease, etc.), unconfirmed neurological symptoms (without MRI or CT), a second malignant tumor, rare types of breast cancer (such as phyllodes, sarcoma, lymphoma, or metastasis), and brain metastases diagnosed within three months of initial diagnosis.

2.3. Definitions

A dictionary of the variables was arranged before gathering patient information, and all data acquisition was supervised and standardized as suggested elsewhere [14]. The expression of ER, PR, HER2, and Ki-67 was determined by immunohistochemistry using the VENTANA BENCHMARK® equipment. The expression of ER and PR were reported as percentages of the reaction's intensity in the cell nuclei, and the H-score and the Allred score were obtained. Samples with percentages $>$ 1% were considered positive [15]. HER2 plasma membrane evaluation was categorized as 0 or 1 + (negative), 2 + (undetermined), and 3 + (positive). If results were undetermined, an in situ hybridization study was conducted [16]. For evaluating Ki-67 expression, the hotspot method was used for reporting the percentage of its value. The modified Scarff–Bloom–Richardson (SBR) grading was used to evaluate the degree of differentiation, nuclear morphology (anisonucleosis), and mitotic activity [17]. The SBR score was divided into SBR I (scores 3–5), SBR II (score 6 or 7), or SBR III (score 8 or 9). Brain metastases (BM) were confirmed with at least one positive MRI or CT. The absence of BM was determined by one or more of the following: lack of neurologic symptoms, negative brain MRI, or having neurologic symptoms with a negative MRI/CT. Tumor size was measured clinically at the first oncologic visit and categorized into $\leq$ 2 cm, $>$ 2 to $\leq$ 5 cm, or $>$ 5 cm. The clinical detection of axillary adenopathy at the first oncologic consultation was cataloged as positive or negative. The anatomic clinical stage (ACS) was divided into Stages I, II, III, or IV using the American Joint Committee on Cancer (AJCC) 7th edition [18] definitions: Stage I if ACS 0, IA, or IB; Stage II if ACS IIA or IIB; Stage III if ACS IIIA, IIIB, or IIIC; and Stage IV if ACS IV.

2.4. Statistical analysis

Patients´characteristics were summarized using descriptive statistics. The cohort was divided into a TD and a VD by a simple division using randomized numbers (1−1−2), as the TRIPOD guidelines[19] advise. Univariate associations were calculated with logistic regression analysis, their odds ratio (OR), and 95% confidence interval (95%CI), and the Wald test P was calculated for each variable. Variables were considered statistically associated with the outcome (the presence or absence of BM) if the P value was < 0.05.

Variable categorization: Continuous variables with a non-normal distribution (age and Ki-67) were entered in the multivariable prediction model with their continuous or categorized value. Model performance was better when age was considered a continuous value and when Ki-67 expression was studied categorically (<10%, 10–19%, 30–40%, and >40%). Categorization of Ki-67 was conducted after generating three knots from their quartile value.

2.5. Model development

Variable selection was based on the available literature on possible predictors, candidate predictors considered for the present model included age at cancer diagnosis, tumor size, axillary node status, clinical stage, ER, PR, HER2, Ki-67, and SBR. The following summarized steps were performed to build the brain metastasis prediction model [12,13]: 1) acquisition of data; 2) identification and management of missing values in the cohort; 3) division of the cohort into a training dataset and a validation dataset; 4) in the TD, associations between variables and the outcome were measured with simple logistic regression analysis; 5) various multivariable prediction models were constructed from the combination of variables significantly associated with the outcome from step 4; 6) the Akaike information criteria (AIC), the C-statistic and the area under the receiver operating curve (AUC) were measured to all generated models; 7) the multivariable model with the lower AIC was chosen and tested with a 10-fold cross-validation; 8) the model's AUC was estimated in the VD; 9) the β coefficients obtained from the multivariable logistic regression analysis were used to generate an equation to determine each individual's risk of having the outcome (brain metastasis risk formula); 10) individual's odds of having BM were calculated with the brain metastasis risk formula; 11) four brain metastasis risk groups were created;12) the hazard ratio was measured for each risk group; and finally, 13) an online formula was generated for a more straightforward clinical use.

2.6. Model performance and internal validation

The AIC was measured for all models, the lower AIC was used as the criteria to choose the better model in the training dataset. The area under the curve (AUC) and C-statistic were used to determine the model's discrimination capacity, which means the ability of the models to predict the patients´risk of brain metastases. A larger AUC indicates a higher discriminative power. To determine the model's internal validation, 10-fold cross-validation was conducted, and the required AUC had to be $\geq$ 0.75 in 9 out of 10 runs to consider a proper internal validation. Regression coefficients were shrunk towards zero to correct for overfitting.

Brain metastasis risk stratification: Once the model was built, individual odds of having BM were calculated using the coefficient values

from the multiple logistic regression analysis (brain metastasis risk calculator) for the whole cohort, the validation, and the training dataset. Patients were divided into four groups (low, medium, high, and very high risk of BM) according to their calculated possibility of having BM. The risk of BM was analyzed for each group using a Cox regression analysis and is presented as hazard ratio (HR), 95%CI, and P value.

2.7. Follow-up

Follow-up was defined as the period from the date of breast cancer diagnosis to the date of death or last visit (if alive). Follow-up was based on electronic health record data, and censoring was considered. The database was last updated on May 31, 2022.

Ethical considerations: The study was approved by the Institution's Scientific and Ethics Research Committees (021/033/SCI and CEI/1556/21).

3. Results

3.1. Study population

A total of 5231 patients were diagnosed with BC between May 2009 and May 2020. Of these patients, 5009 met the eligibility criteria, and 222 did not. Reasons for not including these patients were: absence of complete clinical information n = 86), BM diagnosed during the initial workup or in the first three months since BC diagnosis (n = 44), rare pathologies (n = 43), second malignancy (n = 41), and previous neurologic disease n = 8 (see Fig. 1). In 5.2% (n = 262) of SBR scores and 1.8% (n = 87) of Ki-67 results, missing data were found. These missing values were imputed using the median of each variable's value. After a median follow-up of 53.65 months (IQR 25.56–79.60 months), range 0.1–151.82 months, n = 287 patients (5.7%) were diagnosed with brain metastases.

Variables that were assessed in the study are presented in Table 1. In summary, the median age at BC diagnosis was 51.7 years (IQR, 44.0–60.8); tumor size was ≤ 2 cm in 27%, > 2 but ≤ 5 cm in 46%, and > 5 cm in 26.8%; axillary adenopathy was detected in 54%; and the Stages were I in 12%, II in 47%, III in 27%, and IV in 14%. Tumors were ER+ in 73%, PR+ in 70.8%, and HER2 negative in 79%. Ki-67 expression was < 10% in 19%, 10–19% in 26%, 20%–40 in 33%, and > 40% in 23%. SBR was identified as I in 20%, II in 45%, and III in 35%. The distribution of variables was similar for most of the studied variables between the TD and VD except for tumor size (>2 – 5 cm, P = 0.02, >5 cm P = 0.003) and stage II (P = 0.009).

Table 1
Characteristics of the study population.

Characteristic	Cohort n = 5009	Training dataset n = 3339	Validation dataset n = 1670	P
Median age (IQR), years	51.7 (44.0–60.8)	51.7 (44.1–61.2)	51.6 (43.8–60.1)	0.330
Tumor size, n (%)	1370	919 (27.5%)	451 (27.0%)	0.699
≤ 2 cm	(27.4%)	1582	734 (44.0%)	0.022
> 2–5 cm	2316	(47.4%)	485 (29.0%)	0.003
> 5 cm	(46.2%)	838 (25.1%)		
	1323			
	(26.4%)			
Axillary adenopathy, n (%)	2673	1759	914 (54.7%)	0.170
Present	(53.4%)	(52.7%)	756 (45.3%)	
Absent	2336	1580		
	(46.6%)	(47.3%)		
Anatomic Clinical Stage, n (%)	605 (12.1%)	407 (12.2%)	198 (11.9%)	0.733
I	2396	1641	755 (45.2%)	0.009
II	(47.8%)	(49.1%)	473 (28.3%)	0.106
III	1347	874 (26.2%)	244 (14.6%)	0.036
IV	(26.9%)	417 (12.5%)		
	661 (13.2%)			
Estrogen receptor status, n (%)	3683	2475	1208 (72.3%)	0.176
Positive	(73.5%)	(74.1%)	462 (27.7%)	
Negative	1326	864 (25.9%)		
	(26.5%)			
Progesterone receptor status, n (%)	3546	2378	1168 (69.9%)	0.348
Positive	(70.8%)	(71.2%)	502 (30.1%)	
Negative	1463	961 (28.8%)		
	(29.2%)			
HER2 status, n (%)	1035	667 (20.0%)	368 (22.0%)	0.090
Positive	(20.7%)	2672	1302 (78.0%)	
Negative	3974	(80.0%)		
	(79.3%)			
Ki-67, n (%)	936 (18.7%)	622 (18.6%)	314 (18.8%)	0.882
< 10%	1292	874 (26.2%)	418 (25.0%)	0.382
10–19%	(25.8%)	1082	553 (33.1%)	0.614
20–40%	1635	(32.4%)	385 (23.1%)	0.835
> 40%	(32.6%)	761 (22.8%)		
	1146			
	(22.9%)			
SBR, n (%)	1023 (20.4)	698 (20.9%)	325 (19.5%)	0.232
I (scores 3–5)	2251 (44.9)	1513	738 (44.2%)	0.452
II (score 6 or 7)	1735 (34.6)	(45.3%)	607 (36.3%)	0.072
III (score 8 or 9)		1128		
		(33.8%)		

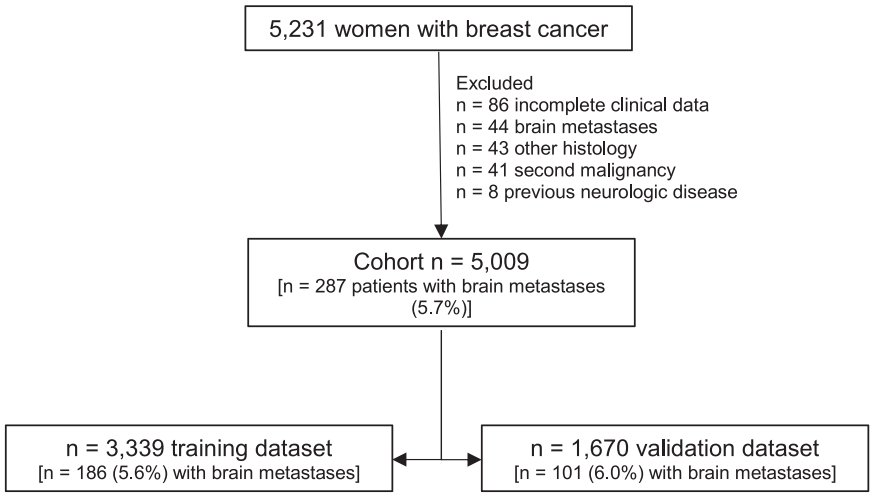

Fig. 1. Schematic representation of the study.

3.2. Model's performance

The cohort (n = 5009) was divided into a TD (n = 3339, 66.6%) and a VD (n = 1670, 33.3%). The variables significantly associated with BM in the simple logistic regression analysis used to generate the prediction model are presented in Table 2, and multivariable associations are shown in Supplemental Table 2. After evaluating the performance of each model, the one that showed better results was built using seven variables. These variables include age, tumor size, axillary adenopathy, clinical stage, ER, Ki-67, and SBR.

The model's performance in the TD exhibited an AUC of 0.79 (95% CI, 0.76–0.82; P < 0.0001) and an AIC of 1241, as shown in Fig. 2A. The model showed suitable consistency in the 10-fold cross-validation, for the model's AUC ranged from 0.792 to 0.800. The model's performance in the VD exhibited an AUC of 0.81 (95%CI, 0.77–0.85; P < 0.0001) and an AIC of 644, as presented in Fig. 2B.

3.3. Brain metastases risk groups

Four brain metastasis risk groups (low, medium, high, and very high risk) were generated according to the brain metastases risk calculation results. Values used for this calculation are presented in Fig. 3. In the VD, the accumulated incidence of brain metastases and the HR for each brain metastasis risk group were the following: in the low-risk group, 1.2% (n = 10/805) were diagnosed with BM; in the medium-risk group, 5.0% (n = 16/321) and HR of 4.01 (95%CI, 1.82 – 8.84; p < 0.0001); in the high-risk group, 8.5% (n = 30/354) and HR of 8.33 (95%CI, 4.07–17.06; P < 0.0001); and in the very high-risk group, 23.7% (n = 45/190) and HR of 29.72 (95%CI, 14.96 – 59.05; P < 0.0001).

In the whole cohort, the accumulated incidence of brain metastases and the HR for each brain metastasis risk group were the following: in the low-risk group, 1.2% (n = 31/2504) were diagnosed with BM; in the medium-risk group, 4.8% (n = 48/999) and HR of 3.95 (95%CI, 2.51 – 6.21; P < 0.0001); in the high-risk group, 9.5% (n = 95/1002) and HR of 9.16 (95%CI, 6.11–13.75; P < 0.0001); and in the very high-risk group, 22.6% (n = 113/500) and HR of 30.75 (95%CI, 20.64 – 45.81; P < 0.0001).

Variable	Units	Value to use in the equation
Age	Years (two decimals)	Age*−0.018
		(*continued on next column*)
		(*continued*)
Variable	Units	Value to use in the equation
Tumor size	≤ 2 cm	0
	> 2–5 cm	0.539
	> 5 cm	1.251
Axillary adenopathy	Absent	0
	Present	0.623
Anatomic clinical stage	I	0
	II	−0.812
	III	−0.427
	IV	0.467
Estrogen receptor status	Positive	0
	Negative	0.640
Ki-67	< 10%	0
	10–19%	0.633
	20–40%	1.008
	> 40%	0.625
SBR	I	0
	II	−0.317
	III	0.044

Table 2
Univariate logistic regression analysis to measure the association of variables with brain metastases that were used to build the prediction model in the training dataset.

Variables	OR (95%CI)	P
Age	0.97 (0.96–0.98)	0.001
Tumor size	Reference	-
≤ 2 cm	1.74 (1.13–2.66)	0.011
> 2–5 cm	7.00 (4.69–10.45)	< 0.0001
> 5 cm		
Axillary adenopathy + vs. −	3.52 (2.63–4.70)	< 0.0001
Anatomic clinical stage	Reference	-
I	1.18 (0.63–2.22)	0.602
II	4.52 (2.47–8.27)	< 0.0001
IIIIV	9.43 (5.13–17.33)	< 0.0001
ER − vs. +	2.75 (2.16–3.50)	< 0.0001
Ki 67	Reference	-
< 10%	2.17 (1.01–4.65)	0.046
10–19%	6.10 (3.05–12.20)	< 0.0001
20–39%	5.93 (2.93–12.05)	< 0.0001
≥ 40%		
SBR	Reference	-
I	1.10 (0.75–1.60)	0.625
II	2.45 (1.71–3.52)	< 0.0001
III		

4. Discussion

In a large cohort of women with BC, a brain metastasis (BM) prediction model constructed with clinical (age, tumor size, axillary adenopathy status, and AJCC ACS) and pathological (ER, Ki-67, and SBR) variables acquired at diagnosis showed a good performance in the TD (AUC = 0.807, P < 0.0001, and AIC = 1350) and in the VD (AUC = 0.827, P < 0.0001, and AIC = 683).

The accumulated incidence of BM in our cohort was 5.7%, similar to what other clinical studies have reported [20–22]. The following variables included in the BM predictive model have previously been associated with a higher risk of BM, and they were replicated in our cohort: a) *Age*: an inverse association between age and the risk of BM has been reported [23,24]. b) *Tumor size*: an increase in size is proportional to the risk of developing BM [8,25,26]. c) *Axillary adenopathy*: its presence increases the risk of having BM [8,21,25,27–29]. d) *ACS*: one of the most valuable tools to guide treatment and a primary prognostic orientation tool in BC; a higher stage has also been associated with a greater risk of BM [20,22,24,30,31]. e) *ER status*: the absence ER expression (ER−) has been associated with a higher risk of metastases [10,23,25,32,33]. f) *Ki-67*: it is a protein used as a proliferation marker in many tumors, including BC; clinically, a higher expression of Ki-67 has been associated with a higher risk of recurrence and shorter overall survival [34,29,35–37]. Here we observed that adding Ki-67 increased the prediction value of our model. The justification for converting Ki-67 into a categorical value was that the model's performance was better when Ki-67 was categorized in quartiles (AIC =1241 when Ki-67 was used as its categorized value vs. AIC =1261 when Ki-37 was used as its continuous value vs. AIC = 1256 when Ki-67 was divided into <20% or ≥20%). There is no universally accepted Ki-67 cutoff value to predict BM. For instance, one study used a cutoff value of < 30% vs. ≥ 30% [34], and another study used > 14% vs. ≤ 14%[36]. g) The *Scarff–Bloom–Richardson (SBR) grading score and its modified version* [17]: they have long been associated with prognosis in BC; SBR has been associated with proliferation, cellular invasion [4,38], and the risk of BM [8,10,21,28,29,31,33,36,39].

HER2 overexpression (HER2 +) defines an aggressive BC subtype. Several studies have reported that HER2 + used as an individual marker is associated with a higher risk of metastatic disease, including BM [10,20,21,27,28,33,36,40,41], and even some authors have proposed the use of prophylactic cranial irradiation in these patients [39]. In the present study, HER2 + was associated with higher odds of BM in the simple regression analysis [OR, 2.27 (95%CI, 1.66–3.11), P < 0.0001] but not in the multivariable analysis [OR, 1.29 (95%CI, 0.91–1.82), P = 0.141]. Adding the variable HER2 to the model did not improve its performance: AIC = 1241 with HER2 vs. 1241 without HER2, and AUC

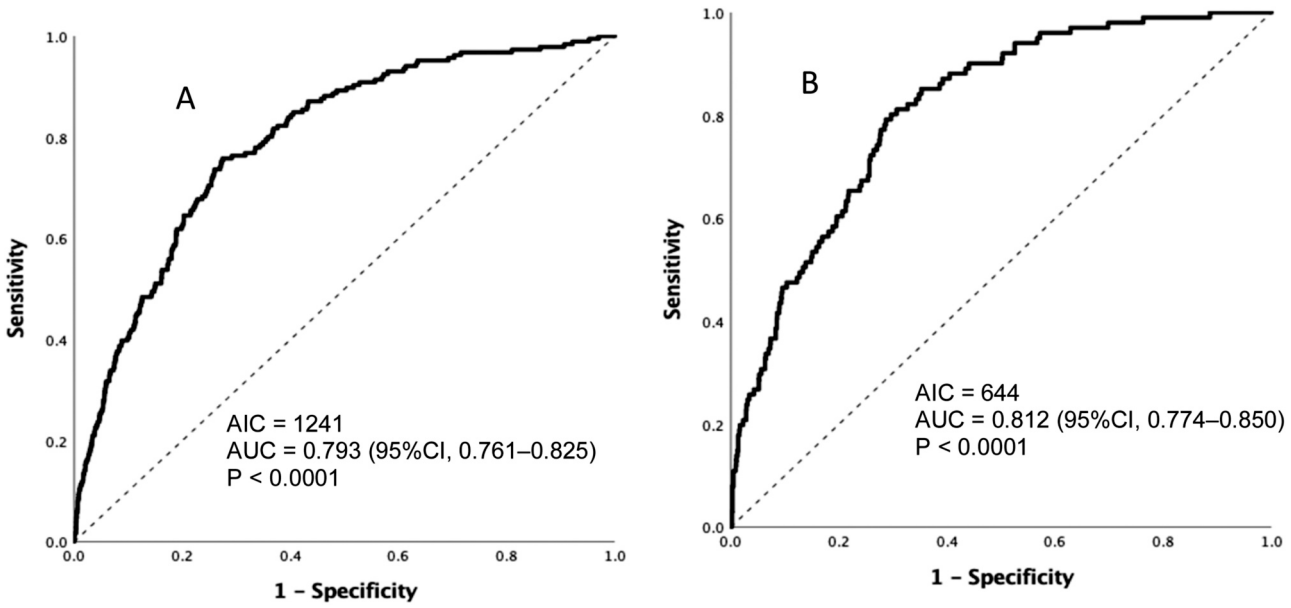

Fig. 2. Brain metastasis prediction model's performance in the training dataset (A) and in the validation dataset (B).

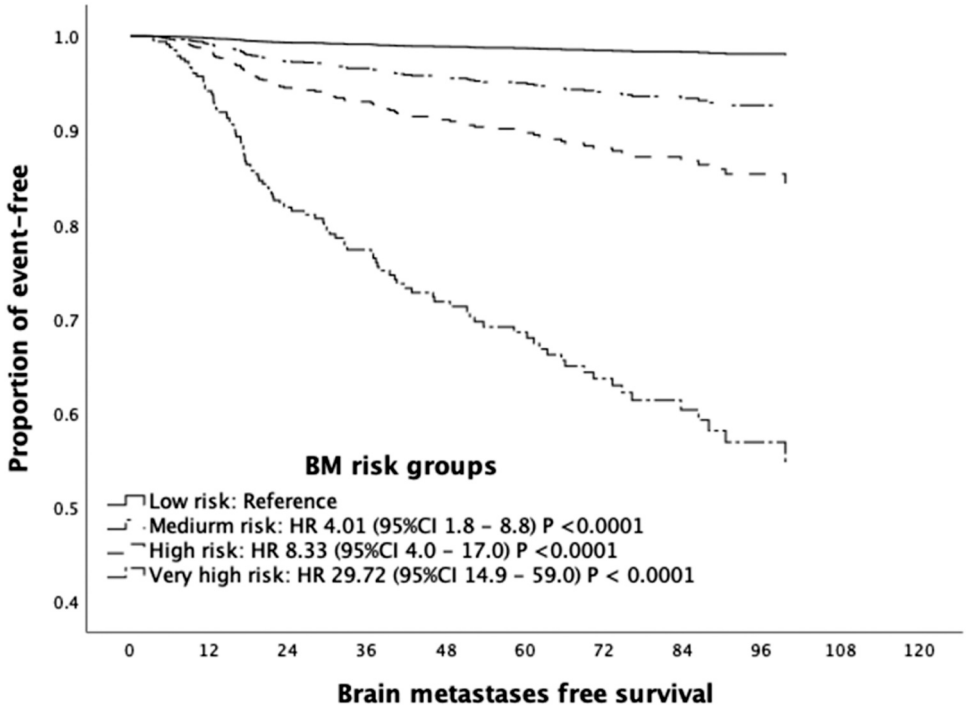

Fig. 3. Cox regression analysis of the risk of brain metastases according to the individual possibility of brain metastases in the validation dataset. Calculated from the equation: P = −3.54 + age + tumor size + axillary adenopathy + clinical stage + ER + Ki-67 + SBR.

= 0.792 vs. 0.793 with HER2.

For the present study, the model was constructed to discriminate patients with a higher possibility of having BM. A valuable tool to assess this discrimination capacity is the calculation of the AUC and C-statistic, calculated from the individual possibility of the outcome. In the present study, the AUC in the TD was 0.793 and that in the VD was 0.812; this indicates that the model that we present has a good discrimination capacity. Calibration of the model, also known as reliability, is used to choose a better model, and the AIC is an often-used tool to measure it. In the present study, the AIC value was 1241 in the TD and 644 in the VD.

There is no unique way to generate a prediction model. Nevertheless, we followed TRIPOD [19] and experts' recommendations [12,13]. A standard error while a model is constructed is to include only variables statistically associated with the outcome. In the eye of an unskilled reader, not seeing the multivariable regression analysis in the results section might be unexpected. Therefore, we included these data as an online table. In the present study, we generated the prediction model with variables associated with the outcome in the simple logistic regression analysis and those that improved the model's performance (i.e., when the variable was included, the model had a lesser AIC and a higher AUC). To ensure the internal validity of the model, the following actions were performed: all database filling and model construction

steps were completed following documented guidelines, a large cohort was analyzed, missing values were handled, division of the cohort into a TD (construction of the model) and a VD (confirmation of the model's performance) was made, and the 10-fold cross-validation was considered.

The strengths of the study were its sample size, the safeguarded steps taken to make sure the model measures what it is meant to measure (good internal validity, calibration, and discrimination capacity), the inclusion of Ki-67 in the prediction model, and the homogeneity of the treatments and population. The limitations of the study were its retrospective nature and the use of data from a single center in a Hispanic population. External validation of the model is mandatory. Thus, we open an invitation to test the model's performance in other centers and demographics.

5. Conclusion

In a retrospective cohort of women with BC, a prediction model built with clinical and pathological variables at diagnosis displayed robust performance to measure the individual odds of BM. If externally validated, this model could help design strategies for screening patients with BC with a higher risk of brain metastases.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Conceptualization: All authors, Methodology: All authors, Validation: **Nancy Reynoso-Noverón**, **Sergio I. Valdés-Ferrer**, and **Antelmo A. Menses-García**, Formal analysis: **Nancy Reynoso-Noverón** and **Bernardo Cacho-Díaz**, Investigation: All authors, Resources: All authors, no funding was received for this work, Data curation: **Nancy Reynoso-Noverón** and **Bernardo Cacho-Díaz**, Supervision: **Nancy Reynoso-Noverón**, Sergio I. **Valdés-Ferrer**, and **Antelmo A. Menses-García**, Writing original draft: **Bernardo Cacho-Díaz**, Writing review and editing: **Nancy Reynoso-Noverón**, **Sergio I. Valdés-Ferrer**, and **Antelmo A. Menses-García**, Visualization: All authors, Project administration: **Nancy Reynoso-Noverón**, Funding acquisition: None.

Declaration of Competing Interest

The authors declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

No funding was received for this work.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.canep.2023.102448.

References

[1] R.L. Siegel, K.D. Miller, H.E. Fuchs, A. Jemal, Cancer statistics, 2022, CA Cancer J. Clin. 72 (1) (2022) 7–33, https://doi.org/10.3322/caac.21708.
[2] N. Klar, M. Rosenzweig, B. Diergaarde, A. Brufsky, Features associated with long-term survival in patients with metastatic breast cancer, Clin. Breast Cancer 19 (4) (2019) 304–310, https://doi.org/10.1016/j.clbc.2019.01.014.
[3] S. Li, J. Zhao, L. Zhu, F. Su, K. Chen, Development and validation of a nomogram predicting the overall survival of stage IV breast cancer patients, Cancer Med. 6 (11) (2017) 2586–2594, https://doi.org/10.1002/cam4.1224.
[4] R. Largillier, J.M. Ferrero, J. Doyen, J. Barriere, M. Namer, V. Mari, et al., Prognostic factors in 1,038 women with metastatic breast cancer, Ann. Oncol. 19 (12) (2008) 2012–2019, https://doi.org/10.1093/annonc/mdn424.
[5] Y. Tsukada, A. Fouad, J.W. Pickren, W.W. Lane, Central nervous system metastasis from breast carcinoma, Autops. Study Cancer 52 (12) (1983) 2349–2354, https://doi.org/10.1002/1097-0142(19831215)52:12<2349::aid-cncr2820521231>3.0.co;2-b.
[6] N. Cifuentes, J.W. Pickren, Metastases from carcinoma of mammary gland: an autopsy study, J. Surg. Oncol. 11 (3) (1979) 193–205, https://doi.org/10.1002/jso.2930110303.
[7] S.M. de la Monte, G.M. Hutchins, G.W. Moore, Influence of age on the metastatic behavior of breast carcinoma, Hum. Pathol. 19 (5) (1988) 529–534, https://doi.org/10.1016/s0046-8177(88)80199-0.
[8] Y.J. Kim, J.S. Kim, I.A. Kim, Molecular subtype predicts incidence and prognosis of brain metastasis from breast cancer in SEER database, J. Cancer Res. Clin. Oncol. 144 (9) (2018) 1803–1816, https://doi.org/10.1007/s00432-018-2697-2.
[9] L. Koniali, A. Hadjisavvas, A. Constantinidou, K. Christodoulou, Y. Christou, C. Demetriou, et al., Risk factors for breast cancer brain metastases: a systematic review, Oncotarget 11 (6) (2020) 650–669, https://doi.org/10.18632/oncotarget.27453.
[10] O. Graesslin, B.S. Abdulkarim, C. Coutant, F. Huguet, Z. Gabos, L. Hsu, et al., Nomogram to predict subsequent brain metastasis in patients with metastatic breast cancer, J. Clin. Oncol. 28 (12) (2010) 2032–2037, https://doi.org/10.1200/JCO.2009.24.6314.
[11] L. Genre, H. Roche, L. Varela, D. Kanoun, M. Ouali, T. Filleron, et al., External validation of a published nomogram for prediction of brain metastasis in patients with extra-cerebral metastatic breast cancer and risk regression analysis, Eur. J. Cancer 72 (2017) 200–209, https://doi.org/10.1016/j.ejca.2016.10.019.
[12] T. Gerds, M. Kattan. Medical Risk Prediction Models: With Ties to Machine Learning, 1st ed..,, Chapman and Hall/CRC,, 2021, https://doi.org/10.1201/9781138384484.
[13] D.W. Hosmer Jr., S. Lemeshow, R.X. Sturdivant. Applied Logistic Regression, 3rd ed..,, John Wiley & Sons, Inc,, 2013, https://doi.org/10.1002/9781118548387.
[14] A.C.M. Jansen, E.S. van Aalst-Cohen, B.A. Hutten, H.R. Büller, J.J.P. Kastelein, M. H. Prins, Guidelines were developed for data collection from medical records for use in retrospective analyses, J. Clin. Epidemiol. 58 (3) (2005) 269–274, https://doi.org/10.1016/j.jclinepi.2004.07.006.
[15] D.C. Allred, R.W. Carlson, D.A. Berry, H.J. Burstein, S.B. Edge, L.J. Goldstein, et al., NCCN task force report: estrogen receptor and progesterone receptor testing in breast cancer by immunohistochemistry, J. Natl. Compr. Canc Netw. 7 (Suppl_6) (2009) S1–S21, https://doi.org/10.6004/jnccn.2009.0079.
[16] A.C. Wolff, M.E.H. Hammond, D.G. Hicks, M. Dowsett, L.M. McShane, K.H. Allison, et al., Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update, J. Clin. Oncol. 31 (31) (2013) 3997–4013, https://doi.org/10.1200/JCO.2013.50.9984.
[17] C.W. Elston, I.O. Ellis, Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up, Histopathology 19 (5) (1991) 403–410, https://doi.org/10.1111/j.1365-2559.1991.tb00229.x.
[18] S.B. Edge, D.R. Byrd, C.C. Compton, A.G. Fritz, F.L. Greene, A. Trotti. AJCC Cancer Staging Manual, 7th ed..,, Springer,, 2011.
[19] K.G.M. Moons, D.G. Altman, J.B. Reitsma, J.P.A. Ioannidis, P. Macaskill, E. W. Steyerberg, et al., Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration, Ann. Intern Med 162 (1) (2015) W1–W73, https://doi.org/10.7326/M14-0698.
[20] M.H. Hung, C.Y. Liu, C.Y. Shiau, C.Y. Hsu, Y.F. Tsai, Y.L. Wang, et al., Effect of age and biological subtype on the risk and timing of brain metastasis in breast cancer patients, PLoS One 9 (2) (2014), e89389, https://doi.org/10.1371/journal.pone.0089389.
[21] H.A. Azim, R. Abdel-Malek, L. Kassem, Predicting brain metastasis in breast cancer patients: stage versus biology, Clin. Breast Cancer 18 (2) (2018) e187–e195, https://doi.org/10.1016/j.clbc.2017.08.004.
[22] J.S. Barnholtz-Sloan, A.E. Sloan, F.G. Davis, F.D. Vigneau, P. Lain, R.E. Sawaya, Incidence proportions of brain metastases in patients diagnosed (1973 to 2001) in the metropolitan detroit cancer surveillance system, J. Clin. Oncol. 22 (14) (2004) 2865–2872, https://doi.org/10.1200/JCO.2004.12.149.
[23] A.M. Martin, D.N. Cagney, P.J. Catalano, L.E. Warren, J.R. Bellon, R.S. Punglia, et al., Brain metastases in newly diagnosed breast cancer: a population-based study, JAMA Oncol. 3 (8) (2017) 1069–1077, https://doi.org/10.1001/jamaoncol.2017.0001.
[24] Q. Shen, A.A. Sahin, K.R. Hess, D. Suki, K.D. Aldape, R. Sawaya, et al., Breast cancer with brain metastases: clinicopathologic features, survival, and paired biomarker analysis, Oncologist 20 (5) (2015) 466–473, https://doi.org/10.1634/theoncologist.2014-0107.
[25] B.C. Pestalozzi, D. Zahrieh, K.N. Price, S.B. Holmberg, J. Lindtner, J. Collins, et al., Identifying breast cancer patients at risk for Central Nervous System (CNS) metastases in trials of the International Breast Cancer Study Group (IBCSG), Ann. Oncol. 17 (6) (2006) 935–944.
[26] Y.L. Tham, K. Sexton, R. Kramer, S. Hilsenbeck, R. Elledge, Primary breast cancer phenotypes associated with propensity for central nervous system metastases, Cancer 107 (4) (2006) 696–704, https://doi.org/10.1093/annonc/mdl064.

[27] F. Heitz, P. Harter, H.J. Lueck, A. Fissler-Eckhoff, F. Lorenz-Salehi, S. Scheil-Bertram, et al., Triple-negative and HER2-overexpressing breast cancers exhibit an elevated risk and an earlier occurrence of cerebral metastases, Eur. J. Cancer 45 (16) (2009) 2792–2798, https://doi.org/10.1016/j.ejca.2009.06.027.
[28] Z. Gabos, R. Sinha, J. Hanson, N. Chauhan, J. Hugh, J.R. Mackey, et al., Prognostic significance of human epidermal growth factor receptor positivity for the development of brain metastasis after newly diagnosed breast cancer, J. Clin. Oncol. 24 (36) (2006) 5658–5663, https://doi.org/10.1200/JCO.2006.07.0250.
[29] M. min Shao, J. Liu, J.S. Vong, Y. Niu, B. Germin, P. Tang, et al., A subset of breast cancer predisposes to brain metastasis, Med Mol. Morphol. 44 (1) (2011) 15–20, https://doi.org/10.1007/s00795-010-0495-2.
[30] F. Heitz, J. Rochon, P. Harter, H.J. Lueck, A. Fisseler-Eckhoff, J. Barinoff, et al., Cerebral metastases in metastatic breast cancer: disease-specific risk factors and survival, Ann. Oncol. 22 (7) (2011) 1571–1581, https://doi.org/10.1093/annonc/mdq625.
[31] N.D. Arvold, K.S. Oh, A. Niemierko, A.G. Taghian, N.U. Lin, R.F. Abi-Raad, et al., Brain metastases after breast-conserving therapy and systemic therapy: incidence and characteristics by biologic subtype, Breast Cancer Res. Treat. 136 (1) (2012) 153–160, https://doi.org/10.1007/s10549-012-2243-x.
[32] V. Rudat, H. El-Sweilmeen, I. Brune-Erber, A.A. Nour, N. Almasri, S. Altuwaijri, et al., Identification of breast cancer patients with a high risk of developing brain metastases: a single-institutional retrospective analysis, BMC Cancer 14 (1) (2014) 289, https://doi.org/10.1186/1471-2407-14-289.
[33] D.G. Hicks, S.M. Short, N.L. Prescott, S.M. Tarr, K.A. Coleman, B.J. Yoder, et al., Breast cancers with brain metastases are more likely to be estrogen receptor negative, express the basal cytokeratin CK5/6, and overexpress HER2 or EGFR, Am. J. Surg. Pathol. 30 (9) (2006) 1097–1104, https://doi.org/10.1097/01.pas.0000213306.05811.b9.
[34] M. Ishihara, H. Mukai, S. Nagai, M. Onozawa, K. Nihei, T. Shimada, et al., Retrospective analysis of risk factors for central nervous system metastases in operable breast cancer: effects of biologic subtype and Ki67 overexpression on survival, Oncology 84 (3) (2013) 135–140, https://doi.org/10.1159/000345321.
[35] A. Martínez-Aranda, V. Hernández, F. Moreno, N. Baixeras, D. Cuadras, A. Urruticoechea, et al., Predictive and prognostic brain metastases assessment in luminal breast cancer patients: FN14 and GRP94 from diagnosis to prophylaxis, Front. Oncol. 7 (2017) 283, https://doi.org/10.3389/fonc.2017.00283.
[36] A.M. Minisini, S. Moroso, L. Gerratana, M. Giangreco, D. Iacono, E. Poletto, et al., Risk factors and survival outcomes in patients with brain metastases from breast cancer, Clin. Exp. Metastas 30 (8) (2013) 951–956, https://doi.org/10.1007/s10585-013-9594-5.
[37] K. Sosińska-Mielcarek, R. Duchnowska, P. Winczura, A. Badzio, H. Majewska, J. Lakomy, et al., Immunohistochemical prediction of brain metastases in patients with advanced breast cancer: The role of Rad51, Breast 22 (6) (2013) 1178–1183, https://doi.org/10.1016/j.breast.2013.08.011.
[38] J.S. Parker, M. Mullins, M.C.U. Cheang, S. Leung, D. Voduc, T. Vickery, et al., Supervised risk predictor of breast cancer based on intrinsic subtypes, J. Clin. Oncol. 27 (8) (2009) 1160–1167, https://doi.org/10.1200/JCO.2008.18.1370.
[39] P. Saip, I. Cicin, Y. Eralp, H. Karagol, S. Kucucuk, R. Cosar Alas, et al., Identification of patients who may benefit from the prophylactic cranial radiotherapy among breast cancer patients with brain metastasis, J. Neurooncol 93 (2) (2009) 243–251, https://doi.org/10.1007/s11060-008-9769-0.
[40] A. Dayan, D. Koca, T. Akman, I. Oztop, H. Ellidokuz, U. Yilmaz, The factors that have an impact on the development of brain metastasis in the patients with breast cancer, J. Cancer Res. Ther. 8 (4) (2012) 542–548, https://doi.org/10.4103/0973-1482.106531.
[41] J. Xue, G. Peng, J.S. Yang, Q. Ding, J. Cheng, Predictive factors of brain metastasis in patients with breast cancer, Med. Oncol. 30 (1) (2013) 337, https://doi.org/10.1007/s12032-012-0337-2.