



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MÉDICAS,
ODONTOLÓGICAS Y DE LA SALUD**

IDENTIFICACIÓN DE MARCADORES ASOCIADOS A METÁSTASIS CEREBRALES EN PACIENTES CON CÁNCER DE MAMA.

Construcción de un modelo predictor de metástasis cerebrales, que logre un adecuado desempeño, a partir de variables clínicas y patológicas obtenidas al diagnóstico de cáncer de mama, en una cohorte retrospectiva de mujeres mexicanas atendidas en el INCAN.

TESIS

**QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN CIENCIAS MÉDICAS**

PRESENTA:

BERNARDO CACHO DÍAZ

TUTOR DE TESIS

DRA. NANCY REYNOSO NOVERÓN
Instituto Nacional de Cancerología

COMITÉ TUTOR

Dr. Antelmo Abelardo Meneses García
Instituto Nacional de Cancerología

Dr. Sergio Iván Valdés Ferrer
Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán

Ciudad Universitaria, CD. MX.
Septiembre 2023



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Dr. David Francisco Cantú de León
Responsable de la Entidad Académica

Dra. Nancy Reynoso Noverón
Tutora

Maestro Bernardo Cacho Díaz
Alumno

Ciudad de México

AGRADECIMIENTOS

INDICE

	PÁGINA
Resumen	5
I. Marco Teórico	8
II. Antecedentes	14
III. Planteamiento del Problema	17
IV. Justificación	17
V. Pregunta de Investigación	17
VI. Objetivos	18
VII. Hipótesis	18
VIII. Diseño del estudio	19
IX. Métodos	23
X. Resultados	27
XI. Discusión	45
XII. Conclusión	57
XIII. Referencias bibliográficas	58

TÍTULO ABREVIADO: Identificación de marcadores asociados a metástasis cerebrales en mujeres con cáncer de mama

TÍTULO EXTENDIDO: Construcción de un modelo predictor de metástasis cerebrales, que logre un adecuado desempeño, a partir de variables clínicas y patológicas obtenidas al diagnóstico de cáncer de mama, en una cohorte retrospectiva de mujeres mexicanas atendidas en el INCAN.

RESUMEN

ANTECEDENTES: El cáncer de mama (CaMa) es la neoplasia con mayor incidencia y mortalidad en las mujeres; además, es la segunda causa de metástasis cerebrales (MeCe). A pesar de los avances en su diagnóstico y tratamiento, la presencia de MeCe es una principal causa de muerte en estas mujeres. Se han descrito, de manera individual, algunos marcadores asociados con la presencia de MeCe y hasta la fecha no existe un modelo predictor de metástasis que tenga un desempeño óptimo. El identificar marcadores clínicos y patológicos asociados a MeCe podría llevar a la construcción de un modelo predictor de MeCe que tenga un adecuado desempeño [determinado como área bajo la curva (AUC) ≥ 0.7], con la finalidad de proponer en un futuro alguna maniobra de detección temprana o terapéutica.

OBJETIVO: Medir la asociación de marcadores clínicos y patológicos al diagnóstico de CaMa con el diagnóstico de MeCe, para identificar variables que permitan construir un modelo predictor de MeCe que logre una adecuada capacidad discriminatoria (AUC 0.7) en una cohorte retrospectiva de mujeres con cáncer de mama atendidas en el INCAN.

HIPÓTESIS: Un modelo predictor de metástasis cerebrales que considere variables clínicas (Edad, tamaño tumoral, adenopatía axilar, etapa clínica) y patológicas (HER2, ER, SBR, Ki-67) obtenidas al diagnóstico de CaMa en una cohorte retrospectiva de mujeres mexicanas atendidas en el INCAN logrará una capacidad de discriminación (AUC ≥ 0.7).

METODOLOGÍA: Tipo de estudio: Cohorte retrospectiva. Desenlace a estudiar: Diagnóstico de metástasis cerebrales. Población de estudio: Mujeres mexicanas con atendidas en el Instituto Nacional de Cancerología con diagnóstico de CaMa realizado entre enero 2009 y mayo 2020. Recolección de datos: Se revisaron los expedientes clínicos electrónicos para obtener datos clínicos y patológicos. Los datos fueron registrados en una base de datos en el programa SPSS v 25.0.

Análisis estadístico: En la cohorte se identificaron e imputaron valores faltantes. Se realizó estadística descriptiva para cada una de las variables estudiadas. Se dividió la cohorte con una división simple (tras la generación de número aleatorios 1-1-2), en una base de entrenamiento (BaseE) y una de validación (BaseV). En la Base E, se midió la asociación entre las variables y el diagnóstico de metástasis con regresión logística simple, se calculó la razón de momios (OR) como medida de asociación, el intervalo de confianza del 95% (IC95%) y la prueba estadística de Wald (P). Las variables que resultaron significativamente asociadas con el desenlace (determinado como aquellas con $P < 0.05$) fueron correlacionadas entre sí. Las variables significativamente asociadas con el desenlace y que no tuvieron correlación entre si fueron empleadas para generar varios modelos predictores. A cada uno de los modelos generados se les calculó su calibración (medida por el criterio de información de Aikake, AIC) y su capacidad discriminatoria (medida por el area bajo la curva, AUC). El modelo con el menor número de variables, menor AIC y mayor AUC fue evaluado con una validación

cruzada (10 Fold-cross validation) para medir su estabilidad y validez interna. Una vez que el modelo demostró un AUC ≥ 0.7 en >9 de las 10 corridas, se midió el desempeño predictor en la Base de validación. Las siguientes fueron determinadas como medidas de desempeño: AIC, AUC y estadístico C. Una vez encontrado el mejor modelo, se obtuvo el valor de β para cada variable (en la cohorte) y se generó la ecuación para medir la posibilidad individual de MeCe. Finalmente, se dividió la posibilidad de MeCe en 4 grupos.

Se utilizaron los criterios de la AJCC para modelos pronósticos y las guías TRIPOD(1,2). Los valores de posibilidad ≤ 0.05 se consideraron como significativos y se usaron pruebas estadísticas de dos colas. Se empleó el programa estadístico SPSS versión 25 para Mac para realizar todos los cálculos.

RESULTADOS: Se revisaron 5481 expedientes de los cuales 5009 cumplieron con los criterios de elegibilidad. La mediana de edad al diagnóstico de CaMa fue de 51.7 (RIC 44-60.8) años, el tamaño de tumor fue ≥ 2 cm en 27%, $>2-5$ en 46% y >5 cm en 26%; se determinó la presencia de adenopatía axilar en 53%; la etapa clínica anatómica fue I 12%, II en 48%, III en 27% y IV en 13%. El 73% presentó sobre expresión de receptores de estógeno y 21% de HER2. La proteína Ki-67 fue determinada como <10 en 19%, 10-19% en 26%, entre 20-40% en 33% y >40 en el 23%. Se incluyeron 2/3 partes de los pacientes en la BaseE (n 3339) y 1/3 en la BaseV (n 1670). Todas las variables estudiadas fueron significativamente asociadas con MeCe.

De los modelos predictores generados, aquel con el menor número de variables, menor AIC y mayor AUC fue el construido a partir de las siguientes variables: Edad (variable continua), tamaño tumoral (TT ≤ 2 cm, 2-5 cm o >5 cm), adenopatía axilar (presente o ausente), etapa clínica anatómica (EC I, II, III o IV), estado de receptores de estrógeno (positivo o negativo), Ki-67 (<10 , 10-19, 20-40 o $\geq 40\%$) y SBR (I, II o III). Este modelo tuvo el siguiente desempeño en la base de entrenamiento AIC = 1241 y AUC = 0.793 (IC95% 0.761-0.825), $P < 0.0001$. En la base de validación el modelo tuvo un AIC = 644 y AUC = 0.812 (IC95% 0.774 – 0.85), $p < 0.0001$. La ecuación de la posibilidad de MeCe fue obtenida y con el valor individual de la posibilidad de MeCe se generaron 4 grupos: Baja posibilidad de MeCe [incidencia acumulada (IA) de MeCe 1.2%], moderada posibilidad [IA = 4.8%, cociente de riesgo (HR) 3.95 (IC95% 2.51 – 6.20), $P < 0.0001$], alta posibilidad [IA = 9.5%, HR 9.15 (IC95% 6.10 – 13.73), $P < 0.0001$] y muy alta posibilidad [IA = 22.6%, HR 30.71 (IC95% 20.61 – 45.76), $P < 0.0001$].

DISCUSION: La incidencia acumulada de MeCe en pacientes con CaMa reportada en la literatura varía entre 0.35-24% en estudios previos y en nuestra cohorte fue del 5.7%. La mediana de SLMeCe en el presente estudio fue de 23.29 meses (IC95% 19.79 – 26.79) similar a lo reportado en otros estudios.

El modelo predictor obtenido fue construido con variables que la literatura ha asociado a MeCe y que se utilizan en la práctica diaria, divididas en a) **variables clínicas:** la edad al diagnóstico de cáncer de mama; se sabe que una menor edad se relaciona con mayor agresividad del tumor, peor pronóstico y mayor prevalencia de MeCe, la etapa clínica anatómica: esta es la variable clínica que más se emplea para tomar decisiones terapéuticas en pacientes con CaMa y debe de incluirse al diagnóstico para también poder establecer el pronóstico, la EC incluye tres principales variables: TNM; T = tamaño tumoral habitualmente categorizada (≤ 2 cm, $>2-5$ cm o >5 cm), N = la presencia de adenopatías siendo el sitio más común de afección ganglionar en la región axilar (presente o ausente) y la presencia de M = metástasis (presente o ausente) y b) **variables patológicas:** Estas variables son habitualmente obtenidas y medidas por el servicio de patología e incluyen: 1) la expresión de receptores de estrógenos (ER), la ausencia de sobreexpresión de receptores de estrógeno (ER-), es la que con mayor frecuencia se encontró en la literatura como asociada a MeCe, 2) Ki-67 es una proteína que sólo se expresa durante la replicación celular y se asocia con un peor pronóstico en CaMa pues se ha asociado a una mayor proliferación celular, se mide como porcentaje

de células que expresan Ki-67, para determinar los niveles de corte en este estudio (<10, 10-19, 20-40 y >40%), se determinaron tres niveles de corte (nudos) de acorde a las percentiles (25, 50 y 75) y 3) el puntaje de Scarff-Bloom-Richardson (SBR), el SBR se utiliza en la clínica para saber el comportamiento del tumor, un mayor SBR se asocia a un comportamiento clínico más agresivo, incluyendo una mayor posibilidad de MeCe; añadir estas dos últimas variables es una aportación del presente trabajo pues existe muy poca literatura que haya estudiado su asociación con MeCe.

El producto final de este proyecto es un modelo predictor de MeCe en mujeres mexicanas, construido con variables previamente asociadas de manera individual con MeCe, con una capacidad discriminativa (AUC) de 0.81 en la base de validación.

CONCLUSIÓN: Un modelo predictor de metástasis cerebrales construido con variables clínicas (edad, tamaño tumoral, presencia de adenopatía axilar y etapa clínica anatómica) y patológicas (ER, Ki-67 y SBR) logró obtener una adecuada capacidad discriminativa y calibración en una cohorte de 5,009 mujeres con cáncer de mama mexicanas atendidas en el INCAN.

I. MARCO TEÓRICO

I.A. CÁNCER DE MAMA.

I.A.1 Prevalencia

El cáncer de mama (CaMa) es la neoplasia maligna más común en mujeres a nivel mundial, representando el 24% de todas las neoplasias sólidas, con 2,088,849 casos y 626,679 muertes registradas en el 2018(3). El CaMa es la primera causa de muerte por neoplasias malignas en mujeres de 25 años y más; los factores de riesgo incluyen los reproductivos, exposición a hormonas, susceptibilidad genética, dietéticos, estilos de vida, ambientales, ocupacionales, personales y exposición a radiación ionizante. La proporción de mujeres que desarrollará CaMa sin factores de riesgo de acorde a la edad es la siguiente: menores de 30 años 1/2,500, a los 40 años 1/200 y a los 50 años 1/50(3).

I.A.2 Clasificación

I.A.2a Etapa clínica:

El CaMa se puede clasificar clínicamente de acorde a la *American Joint Committee of Cancer (AJCC)* séptima Edición 2018, el cual permite emitir un pronóstico y guía el tratamiento. La enfermedad puede ser también dividida como etapa clínica local (EC 0, I y IIa), localmente avanzada (EC \geq IIb - III) o metastásica (EC IV).

I.A.2b Clasificación molecular e inmunofenotipos:

La expresión de ciertos genes y/o proteínas ha llevado a agrupar a los casos de cáncer de mama en grupos(4), también conocidos como inmunofenotipos, que tienen distinto comportamiento clínico, respuesta a tratamiento y pronóstico, de acorde a:

- Receptores hormonales (RH): La expresión de receptores de estrógenos (ER) y progesterona (PR), correlaciona con tumores de bajo grado y mujeres posmenopáusicas. Se presenta en 60% de las pacientes y relaciona de manera inversa con tumores HER2/neu positivos. La expresión de los RH se evalúa con inmunohistoquímica (IHQ), debe ser nuclear y en más del 10% de las células neoplásicas para considerarse positivas.
- HER2/neu (c-erbB-2 o HER2): Se sobreexpresa este receptor en 10-30% de las pacientes con CaMa, por lo general son tumores de alto grado y con RH negativos. El gen HER2 codifica un receptor de tirosina cinasa, y forma parte de la familia de factores de crecimiento epidérmico, relacionados con proliferación, invasión y el pronóstico. Se puede determinar por tres técnicas: IHQ, FISH (fluorescencia), CISH (cromogenia); éstas dos últimas evalúan la amplificación del gen. El resultado debe ser positivo por IHQ y confirmado por FISH o CISH si es 2+; si el resultado por IHQ es 3+ arriba del 90% de las veces se encontrará amplificado.

Los principales inmunofenotipos en los que se divide el CaMa son:

- Luminal A (LumA): Sobre-expresan receptores hormonales (RH), no sobre-expresan HER2 y su valor de Ki67 es menor a 20%
- Luminal B con sobre-expresión de HER2 (LumB/HER2+): Sobre-expresan RH y HER2
- Luminal B sin sobreexpresión de HER2 (LumB(HER2-)): Sobre-expresan RH pero no HER2-neu, Ki 67 \geq 20%.

- HER2 sobre-expresado (HER2+): Sobre-expresan HER2 pero no RH
- Triple negativo (TN): No sobre-expresan RH ni HER2

I.A.2c Clasificación patológica:

Las neoplasias de origen epitelial se denominan luminales, son ER positivos y se subdividen en luminal A y B. Los tumores de origen mioepitelial son ER negativos incluyendo los HER2+ y los triple negativos o de tipo basal no expresan RH ni HER2/neu.

I.B METÁSTASIS CEREBRALES.

I.B.1 Prevalencia de metástasis cerebrales (MeCe) en pacientes con cáncer de mama

Estudios *post-mortem* han documentado que la prevalencia de metástasis cerebrales (MeCe) en CaMa es mayor que la reportada en los ensayos clínicos y práctica médica diaria. En la Tabla I.B.1 se presenta un resumen de la prevalencia de MeCe en los principales estudios *post-mortem* con una media calculada de 29% (rango: 12-65%).

Tabla I.B.1. Frecuencia de MeCe en estudios *post-mortem* en cáncer de mama

Número de pacientes	% MeCe (n calculada)	Referencia
167	28.8% (n 48)	Abrahams et al 1950 (5)
71	24% (n 17)	Lesse et al 1954 (6)
204	25% (n 51)	Aronson et al 1964 (7)
374*	20% <50 años 12% ≥50 años	Viadana et al 1973 (8)
98	17% (n 17)	Posner et al 1978 (9)
707	31% (n 220)	Cifuentes et al 1979 (10)
141	26% (n 37)	Cho et al 1980 (11)
133	30% (n 40)	Hagemester et al 1980 (12)
368	36% (n 133)	Amer et al 1982 (13)
1,044	30% (n 314)	Tsukada et al 1983 (14)
261	18.2% (n 48)	Lamovec et al 1991 (15)
187*	35-65%	De la Monte et al 1988 (16)
Total series = 3,755 Total calculable = 3,194	925 (28.96%)	

*No describen números exactos

I.B.2 Relevancia de la presencia de metástasis cerebrales en cáncer de mama

Varios estudios han demostrado que uno de los principales factores asociados con mortalidad en pacientes con CaMa es la presencia de MeCe. Uno de estos estudios, realizado en 1,015 mujeres con CaMa, reportó la presencia de MeCe en 6.1% (62/1015) con mediana de supervivencia de 6 meses vs 24 para el grupo sin MeCe, encontrando que el principal factor asociado a mortalidad – tras el análisis multivariado – fue la presencia de MeCe (HR 1.17, P <0.0001)(17).

En pacientes con cáncer de mama, el mayor riesgo de muerte asociado a MeCe ha sido confirmado por varios autores(18–22), por ejemplo Kim et al(23) reportó que en

pacientes con EC IV la mediana de supervivencia global fue de 29 meses pero en aquellos con MeCe fue de 11 meses.

I.B.3 Marcadores asociados a metástasis cerebrales en cáncer de mama

En un estudio que incluyó 319 casos de CaMa, se encontró que la sobreexpresión de c-erbB-2 (HER2+) se asoció a una mayor prevalencia MeCe 3/69 (4.3%) vs 1/235 (0.4%) en aquellos HER2-, siendo el primer estudio en reportar la mayor frecuencia de MeCe en este grupo de pacientes (HER2+)(24).

Los estudios que describen la prevalencia de MeCe y las variables asociadas a un mayor riesgo de MeCe en pacientes con CaMa ha sido numerosa desde entonces, los principales estudios que se han publicado se resumen en la tabla I.B.3.

Un análisis de la base de datos SEER (The Surveillance, Epidemiology, and End Results) de los años 2010 al 2014 reportó la prevalencia de MeCe en 206,913 pacientes con CaMa. Según su inmunofenotipo, se encontró una prevalencia de: 0.2% (339/149,861) para pacientes con inmunofenotipo LumA, 0.6% (132/22,648) para aquellas con LumB, 1% (104/9,942) en HER2+, y 0.7% (163/24,462) en pacientes TN. Reportando una prevalencia global de 0.35% (738/206,913) (23). Encontrando como variables asociadas a un incremento en el riesgo de MeCe las siguientes: Edad \leq 40 años, grado patológico \geq 2, tamaño tumoral $>$ 2 cm ($T \geq 2$), ganglios linfáticos regionales con metástasis ipsilaterales a nivel I o II fijos o detectados clínicamente ($N \geq 2$), subtipo LumB, HER2+ y TN, y número de órganos metastásicos ≥ 1 .

La prevalencia de MeCe varía de acorde al inmunofenotipo incluso en pacientes con etapas clínicas tempranas(25), *Arvold et al.* (26) en una cohorte de 1,434 mujeres con CaMa en etapas tempranas (EC I/II) tratadas con cirugía conservadora y quimioterapia adyuvante tras una mediana de seguimiento de 85 meses, encontró que 36/1,434 (2.5%) tuvieron MeCe. La incidencia acumulada de MeCe de acorde al subtipo molecular fue de 0.7% para pacientes con LumA, 12% LumB, 12% HER2+ y 8% para aquellos LumB/HER2+ y 7% en TN; con una mediana de tiempo entre el diagnóstico del tumor primario y metástasis cerebrales (supervivencia libre de metástasis cerebrales, SLMCe) de 51.4 meses [intervalo de confianza (IC) 95% 7.6-107.6 meses]. En este estudio, el 81% de los pacientes con MeCe presentó síntomas neurológicos. Un estudio retrospectivo de 5,529 pacientes con CaMa en el INCAN, reportó 125 casos de MeCe (2.3%) como primer sitio de recurrencia, de las cuales 36/125 (29%) pertenecían al subtipo Luminal A, 36/125 (29%) al triple negativo, 32/125 (26%) al HER2 positivo y 21/125 (17%) Luminal B(21), con una mediana de supervivencia de 14 meses desde el diagnóstico de MeCe. En este proyecto sólo se estudiaron aquellas pacientes que presentaron MeCe como primer sitio de recurrencia.

Las MeCe son más comunes en etapas clínicas (EC) avanzadas, el estudio de *Slimane et al.* reportó la presencia de MeCe en dos series de pacientes con cáncer de mama metastásico (CMM) encontrando una prevalencia de 14.4% (31 de 215 pacientes) en la primera y de 17.2% (199 de 1152 pacientes) de la segunda serie, este estudio reportó como factores de riesgo para MeCe la presencia de CMM, RH negativos, metástasis pulmonares y periodo entre el diagnóstico del tumor primario a primer metástasis $<$ 24 meses(27). Un meta-análisis(28), reportó que la incidencia acumulada de MeCe en pacientes con CMM fue de 31% en el grupo HER2+ (mediana de seguimiento de 30.7 meses), 32% en el grupo TN (mediana de seguimiento de 32.8 meses) y de 15% en los RH+/HER2- (mediana de seguimiento de 33 meses). Con incidencias por paciente/año de 0.13 (95%CI 0.10-0.16) en el subgrupo HER2+, 0.13 (95%CI 0.09-0.2) en el TN y de 0.05 (95%CI 0.03-0.08) para el grupo HR+/HER2-. En

la tabla I.B.3 se resumen las variables asociadas a MeCe y sus medidas de asociación encontradas en la literatura.

Tabla IB3 Prevalencia y asociación de variables con metástasis cerebrales en estudios publicados de pacientes con cáncer de mama

Población	No (%)	Mediana a MeCe	Variables asociadas a MeCe (medidas de asociación)	Ref.
CMM o LA tratados con docetaxel epirubicina	28/92 (30%)	15 meses	-	(29)
CMM	23/155 (15%)	NA	HER2+, número metástasis extracraneales	(30)
CMM	Serie 1 31/215 (14%) Serie 2 199/1152 (17%) Real Serie 2: 20/199 (10%)	13 meses	RH- (HR 4.2 (95%CI 1.7-11) p 0.002, metástasis pulmonares (HR 4.2 95%CI 2-9.3 p 0.0003)	(27)
CMM HER2+ recibieron trastuzumab	23/93 (25%)	9.8 meses (293 días)	Metástasis extracraneales, RH-	(31)
LA o CM inflamatorio	61/768 (8%)	27 meses (2.3 años)	Tamaño tumoral, RH-, grado 3	(32)
CM	Trastuzumab 38/79 (48%) No trastuzumab 123/264 (47%)	NA	Edad, enfermedad ganglionar, metástasis hepáticas	(33)
CMM, <70años, fallecidos	49/219 (22%)	NA	ER-, < 50 años	(34)
HER2+	15/31 (48%) CMM 15/102 (15%)	NA	HER2+	(35)
CMM tratados con epirubicina	124/579 (21%)	NA	DHL, metástasis pulmonares, hepáticas y/o ganglionares, ER-	(36)
HER2+, LA o CMM	4/163 (2%) Lapatinib + capecitabine 11/161 (7%) Capecitabine Total 15/324 (5%)	NA	NA	(37)
CMM	26/1,038 (3%)	NA	MC riesgo de mortalidad (HR 15)	(19)
CMM HER2+	377/1,012	75/1,012 (7.4%) al diagnóstico de CMM Resto 13.3 meses Global 10.8 meses	No trastuzumab o QT, Edad <50 años, RH -, >2 sitios de metástasis	(38)
CM Egipto	160/2,193 (7%)	Seguimiento 5.8 años	Tumor >2 cm (HR 3.6), Enf. ganglionar axillar (HR 4.03), HER2+ (HR 1.89), Grado tumor (HR 1.38)	(39)
CM	RH+/HER2- 361/162,078 (0.22%) 361/6,607 (5.46%) CMM RH+/HER2+ 136/22,376 (0.61%) 136/1,704 (7.98%) CMM RH-/HER2+ 106/9,719 (1.09%) 106/926 (11.45%) CMM TN 173/25,362 (0.68%) 173/1,522 (11.37%) CMM Total 968/238,726 (0.41%) 96/12,801 (7.56%) CMM***	NA	Edad 41-60 (OR 1.41) Edad: 61-80 (OR 1.4), sitios de metástasis extra craneales =2 (OR 1.65), sitios de metástasis extra craneales = 3 (OR 3.4) RH+/HER2+ (OR 1.41), HR-/HER2+ (OR 2.09), TN (OR 2.19)	(18)
CM EC IA-IIIC	125/5,529	29/125 al año (23%) 100/125 (80%) a 3 años	HE2+, TN	(40)
CMM, TN	127/433 (29%)	32/127 al diagnóstico de CMM (25%) Mediana 10 meses	NA	(41)
CM manifestaciones neurológicas	263/970 (27%) 130/1,947 (7%) M0 30/246 (12%) M1	33 meses	Manifestación neurológica (RR 12), Edad <50 (OR 2.5), Lumb/HER2+ (OR 3.1), HER2+ (OR 3.6), TN (OR 2.4), KPS _≤ 70 (OR 6.6), tamaño tumor >2cm (OR	(42)

			2.2), >2 sitios extra craneales de metástasis (OR 3.6) >3 (OR 7.2)	
--	--	--	--	--

*CMM cáncer de mama metastásico, LA Localmente avanzado, NA no aplica, RH receptores hormonales, HR hazard ratio (cociente de riesgos), OR odds ratio (Razón de momios)

**Estudio de tamizaje

*** Subtipo desconocido en 19,191

En resumen: La presencia de metástasis cerebrales se asocia a un pobre pronóstico en pacientes con cáncer de mama y es una condición clínica frecuente. Los marcadores asociados – encontrados en la literatura – con un incremento en la posibilidad de tener el diagnóstico de metástasis cerebrales en pacientes con cáncer de mama son: edad, metástasis sistémicas (extra cerebrales), tamaño de tumor primario >2 cm, grado patológico (SBR), patología ductal, los subtipos HER2+ y triple negativo(43).

I.C. MODELOS PREDICTORES

I.C.1 Utilidad y definiciones

Un modelo predictor es una herramienta que utiliza los valores de los predictores medidos en un paciente a cierto tiempo, para otorgar una probabilidad de presentar un evento (desenlace). El desempeño predictor mide la distancia entre el riesgo predicho y el estado del desenlace (calibración) y la habilidad para estratificar pacientes (discriminación).

Un riesgo (risk) es la probabilidad de que un evento adverso suceda, una oportunidad (chance) es la posibilidad de que un evento positivo suceda. La predicción de un riesgo es una estimación del riesgo basado en una base de datos propósito o estudiada. Un predictor es una variable conocida, en cierto tiempo determinado, que se utiliza para calcular la posibilidad de un desenlace.

Antes de construir o generar un modelo se debe establecer las siguientes definiciones:

- Población de estudio: Describe quienes son elegibles para el uso del modelo e incluye criterios de inclusión/exclusión.
- Tiempo de origen: Describe el tiempo cero o basal, que es el periodo donde debe ser medidas las variables predictoras.
- Blanco de predicción: Describe el evento de interés.
- Riesgos competentes (Competing risks): Describe eventos a partir de los cuales ya no es posible presentar el evento de interés, por ejemplo: Muerte.
- Horizonte de tiempo predictor: Determina el periodo de seguimiento desde el tiempo de origen, por ejemplo: fecha de última cita, seguimiento o muerte.
- Variables predictoras: Son los predictores que se van a medir.

Los métodos para cada una de las etapas de la construcción del modelo incluyen:

- Obtención de datos, llenado y validación de la información de la cohorte.
- Análisis descriptivo.
- Validez interna y división de la cohorte.
- Medidas de asociación simple.
- Medidas de correlación.
- Medidas de asociación múltiple.
- Determinación de variables predictoras.
- Medidas de desempeño
- Validación cruzada.

- Evaluación del desempeño en la base de validación e Interpretación del modelo predictor

Desempeño predictor:

Existen varias maneras de medir el desempeño de un modelo de predicción de riesgo, los más utilizados y descritos en la literatura son el criterio de información de Akaike y el área bajo la curva.

El AIC (Akaike Information Criteria)

Mide la calibración, también conocida como la confiabilidad. Se considera que a menor puntaje mejor calibración. No existe un máximo o un mínimo, sirve para determinar el modelo a escoger y comparar con otros modelos. No existe un método específico y válido para comparar distintos valores de AIC, por lo que en la presente se definió como el mejor AIC a aquel con el menor valor calculado con la siguiente fórmula.

$$AIC = (-2) \log (\text{maximum likelihood}) + 2 (k)$$

k = número de variables predictoras

El AUC (área bajo la curva) y estadístico C (P).

Mide la capacidad de discriminación. La discriminación se refiere a la habilidad del instrumento para estratificar a los sujetos estudiados de acorde a el riesgo de presentar el evento, en este caso metástasis cerebrales. La principal medida de esta discriminación es el área bajo la curva (AUC), también conocido como el estadístico C (de concordancia), un estadístico de concordancia cuantifica la habilidad del modelo para ordenar de manera correcta su riesgo predicho; es decir, el índice de concordancia es la probabilidad que el modelo asigne el riesgo mayor a los sujetos que han tenido el evento de desenlace. Un valor de 100% significa que todos los pacientes fueron correctamente jerarquizados. En general, se considera el valor $\geq 70\%$ (≥ 0.7) para suponer que el modelo predictor tiene un buen índice de concordancia, que es conocido por algunos como la sensibilidad predictora.

Para fines del presente proyecto, se determinó que el desempeño del modelo es mejor con los siguientes criterios:

- 1) Menor número de variables predictoras.
- 2) Menor valor del IC, y
- 3) Mayor AUC.

II. ANTECEDENTES

II.A Modelos predictores de metástasis cerebrales en cáncer de mama

A pesar de los avances en el diagnóstico y tratamiento del cáncer de mama, la presencia de MeCe continúa siendo una condición con alta prevalencia y mortalidad, sin un tratamiento aceptable. Una de las probables causas de la falla terapéutica es quizás la falta de escrutinio (cribado) y, por ende, un diagnóstico tardío; ya que habitualmente las pacientes con MeCe son diagnosticadas ya que tienen síntomas, lo cual es habitual que se deba a una gran carga tumoral cerebral(44).

Aunque existen marcadores (variables aisladas) asociados a un mayor riesgo de MeCe, ninguno puede predecir de manera independiente – con una adecuada sensibilidad predictora – un grupo de pacientes de alto riesgo de desarrollar MeCe, el intento más cercano y diseminado de un modelo predictor de MeCe, es aquel descrito hace más de 10 años por *Graesslin et al.*(45) para mujeres con cáncer de mama metastásico (CMM). Su modelo predictor de metástasis cerebrales (Nomograma) incluyó el análisis de 362 casos de MeCe en 2,136 pacientes con cáncer de mama metastásico (CMM), determinó los siguientes marcadores como variables predictoras de MeCe tras el análisis bi- y multivariado: Edad, el grado patológico >2, el número de metástasis no cerebrales, la sobre expresión de HER2 (independientemente del estado de RH), premenopausia, patología ductal y tamaño tumoral. Este nomograma, ha sido validado en otra cohorte independiente, reportando una AUC de 0.69 (CI95% 0.61-0.77)(46) obtenida en una población francesa. En este estudio encontraron 70 (15.7%) pacientes con metástasis cerebrales en 446 pacientes con cáncer de mama metastásico (CMM). El modelo predictor de *Graesslin et al.* tiene una capacidad predictora <70% – que es el corte mínimo sugerido para un modelo predictor – además, sólo es aplicable para mujeres con cáncer de mama metastásico (Etapa Clínica IV) y fue realizado en una población que incluyó a la raza Hispana en menos del 10% de su muestra.

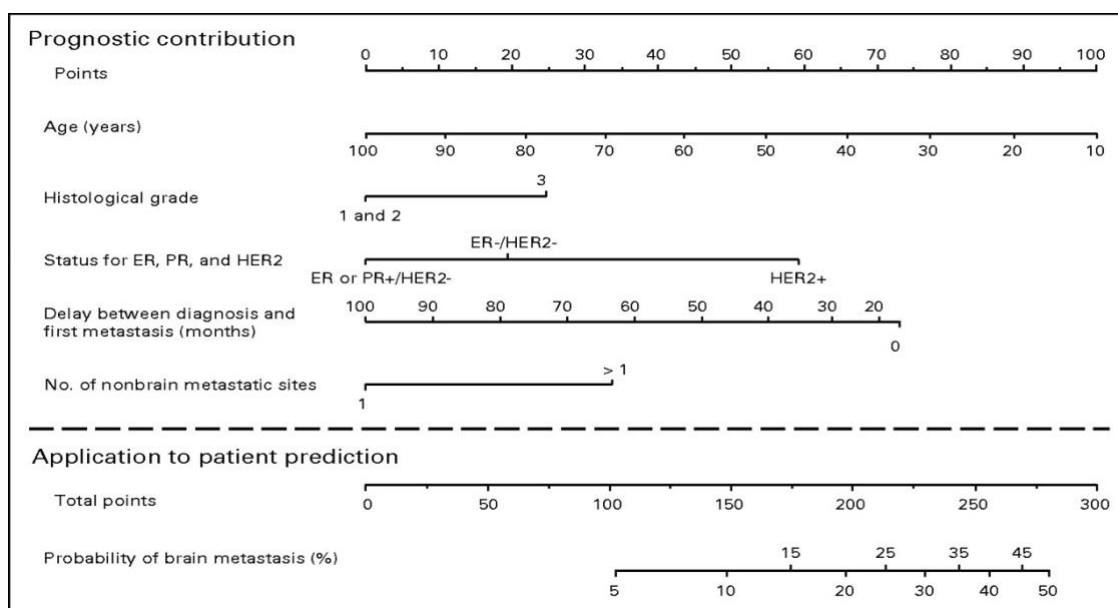


Figura IB4: Nomograma para predecir la probabilidad de metástasis cerebrales en enfermedad metastásica no cerebral. Arriba: El puntaje se obtiene de acuerdo a la contribución de cada parámetro. Abajo: El puntaje obtenido se traduce a la probabilidad de metástasis cerebrales. El puntaje predictor se encuentra en la parte superior para cada variable. El lector o evaluador debe sumar los puntos y el valor de la predicción se encuentra en la parte inferior del nomograma. El total obtenido indica la probabilidad de metástasis cerebrales. El estado de los receptores de estrógeno (ER) y progesterona (PR) se debe determinar por inmunohistoquímica. El estado de HER2 (human epidermal growth factor receptor 2) se determinará por inmunohistoquímica o hibridación fluorescente in situ (FISH).

Hasta la fecha, *NO existe un modelo predictor de Metástasis cerebrales con una adecuada capacidad de discriminación*, por lo tanto, existe un hueco de información que posiblemente pueda ser llenado al generar un nuevo modelo predictor de MeCe que incluya variables clínicas (edad, tamaño tumoral, adenopatía axilar, metástasis extracraneales), de laboratorio (relación neutrófilos/linfocitos) y patológicos (ER, HER2, TN, Ki-67, SBR) logrando obtener un mejor desempeño predictor en pacientes con cáncer de mama.

II.B Utilidad de un modelo pronóstico de metástasis cerebrales en cáncer de mama

Las guías actuales no recomiendan estudios de cribado con resonancia magnética o tomografía computada de encéfalo para determinar qué pacientes con cáncer de mama presentan MeCe(43). Por otro lado, la irradiación profiláctica a cráneo no ha mostrado beneficio en grupos de alto riesgo – CMM y HER2+ – y se ha asociado a efectos adversos serios incluyendo daño cognitivo, por lo que no se recomienda(47–49).

El encontrar un modelo predictor de MeCe con un adecuado desempeño, podría ayudar a identificar un grupo de pacientes que se beneficien de una intervención de detección y/o terapéutica tempranas(50).

II.C Antecedentes directos del presente proyecto

El grupo de personas involucradas en el presente proyecto han publicado los siguientes trabajos en pacientes con cáncer de mama:

Uno de ellos describió la presencia de MeCe como **primer sitio de recurrencia** en 5,529 pacientes con CaMa, encontrando 125 casos con MeCe en el periodo observado del 2007 a 2015(21); no se reportaron marcadores asociados, determinando una prevalencia de MeCe como primer sitio de recurrencia de 2.26%.

En 857 mujeres con cáncer de mama con **manifestaciones neurológicas**, se encontraron metástasis cerebrales en 259 (25%). Tras el análisis multivariado (regresión logística múltiple), el **cuadro clínico** asociado significativamente con la presencia de MeCe incluyó: crisis convulsivas (OR 6.9), alteración visual (OR 2.3), cefalea (OR 2.1) y ataxia (OR 2.1). Otros factores asociados con un incremento en el riesgo de MeCe incluyeron: HER2 positivo (OR 2.2), presencia de metastásis extra-craneal (OR 9.7), metástasis hepáticas (OR 2.6) o metástasis pulmonares (OR 1.7) (44); este estudio estuvo dirigido a mujeres con manifestaciones neurológicas que fueron enviados a un servicio de neuro-oncología y midió la fuerza de asociación entre **manifestaciones clínicas** con la presencia de **MeCe**.

En 970 mujeres con **cáncer de mama y síntomas neurológicos**, 263 (27%) tuvieron MeCe. Los **marcadores asociados al diagnóstico de MeCe** tras análisis multivariado (regresión logística) fueron: Edad < 50 años (OR 2.5), HER2+ (OR 3.6), LumB/HER2 + (OR 3.1), TN (OR 2.4) y KPS ≤ 70 (OR 6.6). Tener alguna manifestación neurológica se asoció con un riesgo relativo de 12 de ser diagnosticada con MeCe. El análisis con regresión logística múltiple para desarrollar MeCe a 3 años determinó los siguientes variables asociadas a MeCe: tamaño de tumor >2 cm (OR 2.2), ≥2 sitios metástasis extra-craneales (OR 4.6) y ≥ 3 sitios (OR 7.2)(42). A partir de este estudio surge la idea de ampliar la población e incluir aquellas mujeres con CaMa con y sin síntomas neurológicos para identificar marcadores asociados a MeCe.

II.D Ki-67 como marcador de MeCe

Una de las posibles causas por las cuales no se ha encontrado un modelo predictor de MeCe que tenga una adecuada capacidad predictora, es el no haber incluido otros marcadores al modelo. Tras una revisión de la literatura, se piensa que agregar Ki-67 como variable a un modelo predictor de MeCe pudiera obtener un buen desempeño.

Ki-67 es una proteína que regula el ciclo celular, se sobre-expresa durante la mitosis y está presente en todas las células que proliferan. Desde hace varios años, su sobre-expresión, medida generalmente por inmunohistoquímica con el anticuerpo MIB-1, se emplea como marcador de proliferación en cáncer de mama y varios estudios han asociado Ki-67 con un mayor riesgo de recurrencia, una disminución en la supervivencia global y como predictor de respuesta al tratamiento sistémico(51–53).

Un estudio retrospectivo en 591 pacientes con CaMa en etapas clínicas I-III, reportó que tras el análisis multivariable (AMV) un Ki-67 $\geq 30\%$ vs $<30\%$ se asoció con un mayor riesgo de MeCe HR 3.9 (95%CI 1.2-12.9) P = 0.026 y una supervivencia libre de metástasis cerebrales más breve ($<30\%$ = 12.3 meses vs $\geq 30\%$ 23.9 meses)(54). El valor de corte para el valor de Ki67 aún no está bien definido, Jung et al, determinó que la expresión en $\geq 10\%$ de las células tumorales puede ser utilizado como nivel de corte para estimar la supervivencia global y enfermedad libre de recurrencia. Jalava et al.(55) dividió a los pacientes en tres grupos de acuerdo al porcentaje de expresión de Ki-67 en ≤ 15 , 15-30 y $>30\%$; de hecho, el consenso de expertos de St. Gallen en 2009(56) aprobó este nivel de corte para Ki-67 para pacientes con CaMa; estos valores de corte han sido establecidos como marcadores pronósticos en pacientes con CaMa. El nivel de Ki-67, junto con la expresión de ER, PR y HER2 ha demostrado ser un buen indicador de pronóstico de pacientes con CaMa. El índice denominado IHC4(57) propuso que el porcentaje de expresión de Ki-67 es uno de los criterios para diferenciar entre el inmunofenotipo LumA y LumB/HER2- cuya única diferencia para el primero es tener expresión de Ki67 <20 y el segundo $\geq 20\%$ como se describe en la sección I.A.2b.

Un estudio confirmó que la expresión de Ki-67 se asoció con un mayor desarrollo de metástasis, concluyendo que una mayor expresión de Ki-67 puede ser un indicador de progresión del ciclo celular y está incrementada en aquellos pacientes con enfermedad metastásica temprana en pacientes con CaMa (58).

La proteína Ki-67 fue originalmente descrita como un anticuerpo monoclonal de 359 kDa, obtenido al inmunizar ratones con el núcleo de líneas celulares de linfoma de Hodgkin en 1991(59). El nombre Ki-67 fue definido por primera vez por la ciudad de origen (Kiel) y el número original de la clona (67). Ki-67 es una proteína de unión al ácido desoxirribonucleico nuclear (ADN) que se expresa en todos los animales vertebrados y es usado como marcador de proliferación celular. Varios reportes del análisis del ciclo celular han demostrado la presencia de Ki-67 en las fases G₁, S y G₂ del ciclo celular pero no en la fase quiescente G₀. En diversas enfermedades oncológicas, un índice elevado de Ki-67 ha sido asociado con una peor evolución clínica(60), por lo que es ampliamente utilizado como marcador de pronóstico en diversas neoplasias. Ki-67 es una proteína nuclear y nucleolar codificada en el gene MKI.67 localizado en el cromosoma 10q26.2 en humanos(60).

En pacientes con CaMa, se ha demostrado que la expresión de Ki-67 está asociado con el pronóstico, demostrando una asociación significativa entre su expresión y el riesgo de recurrencia (local o distante) y mortalidad, así como con la supervivencia global y supervivencia libre de enfermedad(52).

III. PLANTEAMIENTO DEL PROBLEMA

El cáncer de mama es la neoplasia más frecuente y con mayor mortalidad en mujeres, la principal causa de muerte en mujeres con cáncer de mama es la enfermedad metastásica, siendo las metástasis cerebrales las que han sido asociadas con mortalidad. Se han identificado algunos marcadores aislados asociados con un mayor riesgo de metástasis cerebrales como la edad, estado de receptores de estrógenos, HER2, SBR y metástasis extra-craneales; sin embargo, no se ha logrado construir un modelo predictor que tenga un adecuado desempeño que permita identificar un grupo de alto riesgo de metástasis cerebrales.

IV. JUSTIFICACIÓN

El CaMa es la neoplasia más frecuente en las mujeres y es la segunda causa de MeCe en el mundo. A pesar de los avances en su diagnóstico y tratamiento, la presencia de metástasis es principal causa de muerte en mujeres con CaMa y de éstas, las metástasis cerebrales son las que se han asociado a una menor supervivencia global. Se han descrito algunos marcadores asociados con MeCe por ejemplo, edad al diagnóstico de CaMa, tamaño tumoral, adenopatía axilar, etapa clínica, ER, HER2, SBR; no obstante, falta aún explorar algunos marcadores como el Ki-67. Hasta la fecha, no existe un modelo predictor de metástasis cerebrales que tenga un buen desempeño predictor y no existe un modelo predictor de MeCe construido en población mexicana o en mujeres con CaMa en cualquier etapa clínica. El identificar marcadores clínicos o patológicos asociados a MeCe podría ayudar a generar un modelo predictor, que tenga un buen desempeño, con la finalidad de proponer en un futuro, una intervención de tamizaje, diagnóstica o terapéutica

V. PREGUNTA DE INVESTIGACIÓN

¿Un modelo predictor de metástasis cerebrales, construido a partir de marcadores clínicos y patológicos al diagnóstico de cáncer de mama, en una cohorte retrospectiva de mujeres mexicanas atendidas en el INCAN, puede lograr una adecuada capacidad discriminatoria ($AUC \geq 0.7$)?

VI. OBJETIVOS

V.1 Objetivo general

Medir la asociación de marcadores clínicos y patológicos al diagnóstico de cáncer de mama con el diagnóstico de metástasis cerebrales, para identificar variables que permitan construir un modelo predictor de metástasis cerebrales que logre una adecuada capacidad discriminatoria ($AUC \geq 0.70$) en una cohorte retrospectiva de mujeres mexicanas con cáncer de mama atendidas en el INCAN.

V.2 Objetivos específicos

1. Describir las variables clínicas [Edad, etapa clínica anatómica (TNM)] y patológicas (HER2, ER, SBR, Ki-67) en la cohorte.
2. Medir la asociación entre las variables clínicas y patológicas con el desenlace (MeCe) en la base de entrenamiento (BaseE).
3. Generar modelos predictores con las variables asociadas a MeCe (BaseE).
4. Medir el desempeño [calibración (AIC) y capacidad predictora (AUC)] de cada modelo generado (BaseE).
5. Determinar el modelo a estudiar con los criterios: menor cantidad de variables, menor AIC y mayor AUC (BaseE).
6. Realizar la validación cruzada del modelo (BaseE).
7. Medir el desempeño del modelo predictor en la base de validación.
8. Establecer la interpretación clínica del modelo predictor.

VII. HIPÓTESIS

Un modelo predictor de metástasis cerebrales que considere variables clínicas [Edad, tamaño tumoral, adenopatía axilar, etapa clínica anatómica (TNM)] y patológicas (HER2, ER, SBR, Ki-67) obtenidas al diagnóstico de cáncer de mama en una cohorte de mujeres mexicanas atendidas en el INCAN logrará una capacidad discriminatoria ≥ 0.7 .

VIII. DISEÑO DEL ESTUDIO

Se siguieron las guías TRIPOD(1,2) para el reporte de un modelo multivariable predictor, los pasos recomendados en estas guías se describen a continuación:

VIII.A Tipo de estudio

Cohorte retrospectiva. Los estudios de cohorte son el mejor método para el desarrollo de modelos de predicción(61). En realidad es un estudio híbrido de casos y controles anidado en una cohorte(62) (Szklo Moyses and Nieto F. Javier. Chapter 1. Basic study designs in analytical epidemiology. EPIDEMIOLOGY: Beyond the Basics. Aspen Publishers, Inc. 1st Ed. Pp 3-51. Pages 495), en este tipo de estudio se ha demostrado que la posibilidad del evento (odds) es un estimado sin sesgo del riesgo de presentar el desenlace.

VIII.B Población de estudio

Mujeres con cáncer de mama atendidas en el Instituto Nacional de Cancerología con diagnóstico patológico de CaMa realizado entre enero 2009 a mayo 2020.

VIII.B.1 Cálculo del tamaño de la muestra

Cuando cada sujeto en la población recibe un riesgo predicho distinto a partir del modelo, el blanco del modelo predictivo no es una prueba de hipótesis única en la que aplique una prueba de poder estadístico estándar. Una regla de pulgar es incluir al menos 10 eventos por cada predictor a estudiar. Para fines del presente estudio, la fórmula empleada para el cálculo del tamaño de la muestra es la determinada por van Smeden M *et al*(63)

$$N = 10 k / p$$

Donde k es el número de variables que se incluirán en el modelo. En el presente estudio se han considerado las siguientes variables: Edad, ER, HER2, Triple negativo, SBR, Ki67, NLR, tamaño tumoral, presencia de adenopatía axilar = 10 variables.

La proporción de metástasis cerebrales en estudios previos fue de 0.11. Se agregó un 20% de error, por lo tanto: $n=10*10/0.11 = 909 + 20\% = 1,091$ mujeres con cáncer de mama fueron calculadas para generar un modelo predictor de metástasis cerebrales.

VIII.C Recolección de datos

Se pidieron los números de expedientes de pacientes con cáncer de mama atendidos en el INCAN de enero 2009 a mayo 2020 al departamento de archivo clínico. El criterio de búsqueda fue: "cáncer de mama" y/o "tumor de mama". El investigador principal y su equipo de médicos pasantes del servicio social en medicina (MPSS) revisaron los expedientes clínicos para obtener datos clínicos y patológicos. Los datos fueron registrados en una base de datos en el programa SPSS v 25.0.

VIII.D Criterios de elegibilidad

Criterios de inclusión

- Mujeres con edad ≥ 18 años.
- Con diagnóstico de cáncer de mama confirmado por el servicio de patología del INCAN entre enero 2009 y mayo 2020.

- Contar con descripción de edad, tamaño tumoral (T), el estado ganglionar (N) y etapa clínica al diagnóstico.
- Contar con resultado de ER, HER2
- Seguimiento en el INCAN por al menos 1 mes.

Criterios de exclusión

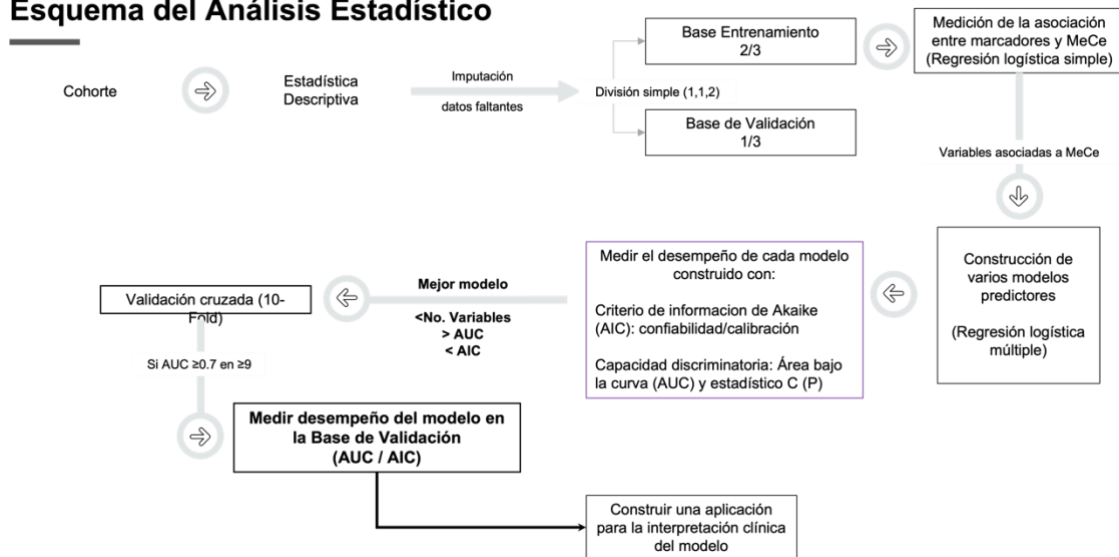
- Presencia de un segundo tumor primario sincrónico.
- Diagnóstico de metástasis cerebrales en los primeros 3 meses desde el diagnóstico de cáncer de mama.
- Enfermedad neurológica previa al diagnóstico de cáncer de mama (por ejemplo, epilepsia, cefalea primaria, retraso psicomotor, neuro-infección).
- Presencia de síntomas neurológicos sin confirmación diagnóstica.
- Presencia de síntomas neurológicos y alguna condición médica que proscriba la realización de resonancia magnética o tomografía computada.
- Cáncer de mama bilateral
- Patología compatible con tumor Phyllodes, sarcoma, papilar, linfoma o metástasis/invasión de otro tumor primario.

Criterios de eliminación

- No hay criterios de eliminación, los pacientes perdidos al seguimiento serán considerados como censuras en el análisis de metástasis cerebrales y supervivencia libre de MeCe.

El esquema general del análisis estadístico para el estudio es el siguiente:

Esquema del Análisis Estadístico



VIII.E Variables y definiciones operacionales:

Metástasis cerebrales (MeCe): Presente (Código = 1): Diagnóstico confirmado con: a) estudio de imagen (Resonancia Magnética y/o tomografía de cráneo) de actividad tumoral en encéfalo o meníngeo, y/o b) estudio de líquido cefalorraquídeo con citología positiva para meningitis neoplásica. No presente (Código = 0): Sin diagnóstico establecido en el expediente clínico de MeCe y/o con estudio de imagen negativo y/o estudio de LCR negativo. Esta es la variable de desenlace.

Fecha de nacimiento: Formato dd.mm.aaaa, día del calendario natural del nacimiento.

Fecha de diagnóstico de cáncer de mama: Formato dd.mm.aaaa, día del calendario natural cuando se confirmó patológicamente el diagnóstico de cáncer de mama.

Fecha de diagnóstico de metástasis cerebrales: Formato dd.mm.aaaa, día del calendario natural cuando se le realizó estudio de imagen (Tomografía o Resonancia Magnética) o LCR que confirma la presencia de metástasis cerebrales.

Fecha de muerte: Formato dd.mm.aaaa, día del calendario natural del fallecimiento registrado en el expediente clínico electrónico.

Fecha de última consulta: Formato dd.mm.aaaa día del calendario natural de la última consulta a la que acudió al INCAN o se tuvo contacto por vía telefónica de acorde al expediente clínico electrónico.

Muerte: Fallecimiento 1, censurado 0

Tabla VII.E.1 Etapa clínica: Obtenido al diagnóstico de cáncer de mama, de acorde al AJCC Cancer Staging Manual, octava edición.

Etapa Clínica	Etapa Clínica Anatómica	TNM
	0	Tis, N0, M0
I	IA	T1, N0, M0
	IB	T0, N1mi, M0 T1, N1mi, M0
II	IIA	T0, N1, M0 T1, N1, M0 T2, N0, M0
	IIB	T2, N1, M0 T3, N0, M0
III	III A	T0, N2, M0 T1, N2, M0 T2, N2, M0 T3, N1, M0 T3, N2, M0
	III B	T4, N0, M0 T4, N1, M0 T4, N2, M0
	III C	Cualquier T, N3, M0
IV	IV	Cualquier T, cualquier N, M1

Tamaño de tumor (T): Se definirá de acorde al diámetro máximo determinado por exploración física establecida en el expediente electrónico, las categorías incluyen: T1 si el diámetro máximo ≤ 2 cm, T2 si >2 pero ≤ 5 cm y T3 si > 5 cm o afección de piel.

Estado de los receptores de estrógeno (ER): Medidos como positivos (1), o negativos (0) de acorde al informe de patología del INCAN. Se considerará en caso de duda o no estar el resultado como positivo si H-Score es $\geq 10\%$ y/o Allred-Score ≥ 3 .

Estado de los receptores de progesterona (PR): Medidos como positivos (1), o negativos (0) de acorde al informe de patología del INCAN. Se considerará en caso de duda o no estar el resultado como positivo si H-Score es $\geq 10\%$ y/o Allred-Score ≥ 3 .

Estado del HER2 (HER2): Medidos como positivos (1), o negativos (0) de acorde al informe de patología del INCAN. Se considerará como positivo en caso de tener resultado por inmunohistoquímica 3+ y/o FISH positivo.

Estado del Ki-67 (Ki-67): Medido en porcentaje de acorde al informe patológico del INCAN.

Grado patológico (SBR): Medido con la escala de Scarff-Bloom-Richardson (SBR)(64,65) medido del 3 al 9; se agrupó como grado I (SBR 3-5), grado II (SBR 6,7) o grado III (SBR 8,9) de acorde al informe patológico del INCAN.

Adenopatía axilar (Enfermedad ganglionar axilar): Presencia de nódulos o ganglios con actividad tumoral clínicamente detectados al momento del diagnóstico de cáncer (1) o ausencia de estos (0).

IX. MÉTODOS

IX.a Obtención de datos, llenado y validación de la información de la cohorte

Utilizando las definiciones previamente detalladas, se capturaron los datos obtenidos del expediente clínico electrónico en la computadora localizada en el área de neurociencias por el MPSS y el alumno. Se hizo una prueba de adiestramiento del llenado y supervisión de la base de datos con los primeros 300 expedientes revisados. Durante el periodo de llenado de la base de datos, se repasaron las definiciones operacionales de manera mensual bajo supervisión de dos expertos en cáncer de mama.

Cada semana, se realizó un corte en el llenado de datos y se realizaron pruebas estadísticas para buscar errores de llenado, errores de teclado y datos atípicos. En caso de duda sobre algún paciente o variable específica, cuyo valor no estuviera definido en el diccionario de variables, se solicitó la intervención de uno de los expertos en el tema. Los expertos también supervisaron que la base de datos contuviera veracidad de la información al revisar aleatoriamente la información de la base de datos y corroborando su contenido con el expediente clínico electrónico para aproximadamente el 5% de los expedientes.

La captura de los datos se llevó a cabo por el MPSS y el alumno. La revisión se llevó a cabo por el alumno y sus tutores de tesis. La *validación* se llevó a cabo por los expertos en el tema. Todos los datos se encontraron estandarizados y definidos antes del inicio de captura de la base de datos.

La calidad de la colección de los datos se llevó a cabo siguiendo los pasos establecidos en las guías publicadas para la colección de datos desde los expedientes clínicos, considerando los riesgos y medidas para reducir inconsistencias que se resumen en la Tabla IX.a.

Tabla IX.a. Guías para la recolección de datos desde los expedientes clínicos(66).

Riesgos	Estrategias para disminuir inconsistencias
Flujo de información del paciente al expediente clínico	
Información incorrecta proporcionada por el paciente y/o registrada de manera incorrecta	Verificar la información en la documentación original y múltiples notas médicas Realizar cuestionarios adicionales o re interrogar a un grupo seleccionado de pacientes
Reportes diagnósticos de un paciente equivocado	Corroborar que los reportes correspondan al paciente
Flujo de información del expediente clínico a la base de datos	
Expediente clínico Los datos no están accesibles El expediente está incompleto	Revisar acceso con un estudio piloto, envío de cuestionarios o re-interrogar a los pacientes Cooperación cercana con médicos tratantes quienes podrían complementar información
Colectores de información Interpretación incorrecta de los datos Colección no estandarizada de los datos Colección no autenticada de datos Desenlace incierto Escritura ilegible Sesgo de los colectores Falta de información de los colectores Transferencia equivocada de los datos desde el expediente a la base de datos	Los recolectores deben tener la información y formación médica Consultar con expertos en el tema sobre las definiciones y desenlaces Crear un diccionario de definiciones, manual de procedimientos y consenso de criterios Crear un formato de recolección Revisar continuamente que la información de la base de datos sea lo que está descrito en los expedientes clínicos por los colectores y un tercero Utilizar solo un equipo de recolectores en todo el proceso. Realizar estudio inter-observador para evaluar variabilidad.

	Cegado de desenlaces al recolector Organizar reuniones frecuentes para reafirmar definiciones y criterios Realizar verificación aleatoria de la información Realizar revisión de la consistencia de la información y búsqueda de datos atípicos
--	--

IX.b. Valores faltantes (perdidos)

Aquellos valores no encontrados por falta de medición o cualquier otra razón deben ser siempre atendidos. Habitualmente, se sugiere no dejar “vacíos” estos valores para poder calcular medidas de asociación como la razón de momios o razón de riesgos. Los métodos recomendados empleados para llenar este vacío son la imputación múltiple y la probabilidad ponderada inversa, pero antes de dar un valor se debe asegurar que los valores faltantes se presenten de manera completamente aleatoria y que éstos como grupo no tengan relación con el desenlace.

En este estudio los valores faltantes fueron encontrados en menos del 10% (en 1.8% de los expedientes no se encontró la determinación de Ki-67 y en 5.2% de SBR). Los datos faltantes fueron atendidos con imputación múltiple de acorde a la distribución de la variable, al tener una distribución “no-normal” fueron imputados con la mediana del valor de la variable (Gerds TA, Kattan MW. Chapter 7 What would me make an expert?. Medical Risk Prediction Models. With Ties to Machine Learning. 1st Edition. CRC Press. Pp 180-223).

IX.c Estadística descriptiva

Se utilizó para las variables cuantitativas con distribución normal medidas de tendencia central (media) y medidas de dispersión (desviación estándar, DE); para aquellas con distribución no normal: medianas y rango inter-cuartil (RIC). Para las variables nominales y ordinales se usaron números y porcentajes. La distribución de las variables cuantitativas se consideró, no-normal para las variables edad, Ki-67 y SBR, el resto de las variables se consideraron como cualitativas nominales (categóricas).

IX.d División de la cohorte y validez interna

En teoría, lo que realmente queremos saber al generar un modelo predictor, es que tan bien el modelo predecirá el riesgo al ser aplicado en pacientes vistos a futuro. Una de las principales limitantes para el desarrollo de un modelo es el acceso limitado a información de la población blanco. Al poder contar con una chorte numerosa, se dividió la cohorte en una base de entrenamiento y una base de validación a través de una división simple tras generar números aleatorios 1-1-2.

Al realizar una división simple de la base de datos, se “esconde” parte de los datos de todos los pasos utilizados en la construcción del modelo para luego medir el desempeño condicionado del modelo final en los datos no incluidos en la etapa de construcción del modelo. Éste es el método de elección para la construcción de un nuevo modelo predictor. La regla de oro de la division es dividir la base total (base “t”) de manera aleatoria para incluir 63% de los datos en la base de entrenamiento (base “e”) y el resto de los datos en la base de validación (base “v” = base t – base e). Hacer esta división permite al investigador saber que está midiendo en realidad lo que quiere medir, en este caso la posibilidad de ser diagnosticada con metástasis cerebrales.

Existen múltiples métodos para dividir una cohorte(67,68), en una base de entrenamiento (BaseE) y una de validación (BaseV). La revisión de la literatura realizada apoya con mayor evidencia que el mejor método para realizar la división de la cohorte es con una división simple no aleatoria(1), tal y como fue hecho en este proyecto.

Base de entrenamiento

La base de entrenamiento, conocida también como base de aprendizaje o base de prueba, se utilizó para:

- a) Medir la asociación entre las variables candidatas a ser incluidas en el modelo predictor con el desenlace (MeCe). Esta asociación fue medida con una regresión logística simple. Se determinó a la asociación como significativa si el valor de la prueba estadística de Wald (P) fue <0.05 .
- b) Medir la correlación entre estas variables significativamente asociadas con metástasis cerebrales para Identificar a aquellas que tuvieran correlación entre sí.
- c) Generar varios modelos predictores con regresión logística múltiple con combinaciones de las variables encontradas en a) que no tuvieran correlación entre si.
- d) Medir el desempeño predictor del modelo con el criterio de información de Akaike (AIC), el área bajo la curva (AUC) y la prueba estadística C (P).
- e) Identificar el modelo con el mejor desempeño, determinado como aquel con menor número de variables, menor AIC y mayor AUC.
- f) Realizar la validación cruzada x 10 (10-fold cross validation) para evaluar la validez interna (estabilidad) del modelo.

Base de validación

La base de validación sirvió para validar el desempeño predictor del modelo elegido en la base de entrenamiento.

La literatura menciona que mayor es el número de datos en la base de entrenamiento, el modelo tendrá un mejor desempeño y a mayor número de datos que se incluyen en la base de validación, el desempeño del modelo será más cercano al desempeño con datos desconocidos. El dividir la cohorte en una base de entrenamiento y una de prueba nos permite obtener la certeza de que se está midiendo lo que realmente se quiere medir, también conocida como validez interna.

IX.e Medidas de correlación e independencia de las variables

Para demostrar que cada variable está relacionada de forma independiente con el desenlace se realizó medida de correlación de Pearson o Spearman para las variables cuantitativas; para las variables cualitativas se midió la correlación con las fórmulas de Cochran-Mantel-Haenszel (CMH) o $= [(n-1)/n]X^2$; donde X^2 es el estadístico de Pearson o la fórmula de Mantel $= (n-1)r^2$. Las variables con un valor >0.7 fueron excluidas.

(Agresti A. Chapter 3. Inference for two-way contingency tables. *Categorical Data Analysis*. Third Edition. Wiley. 201. Pp 69-113 y Agresti A. Chapter 8. Models for multinomial responses. *Categorical Data Analysis*. Third Edition. Wiley. 201. Pp 293-338)

IX.f Desempeño predictor

El desempeño predictor mide la distancia entre el riesgo predicho y el estado de desenlace (calibración = AIC) y la capacidad para jerarquizar pacientes (discriminación = AUC). En el caso del presente estudio, el evento que se quiere predecir es el diagnóstico de metástasis cerebrales.

Los criterios utilizados para medir el desempeño del modelo predictor fueron:

- El criterio de información de Akaike (AIC): Mide la distancia entre el riesgo predicho y el estado de desenlace. También es conocida como medida de calibración o confianza.
- El área bajo la curva (AUC) el estadístico C: La capacidad de discriminación fue medida por el AUC, su intervalo de confianza al 95% (IC95%) y el estadístico C.

IX.g Validación cruzada

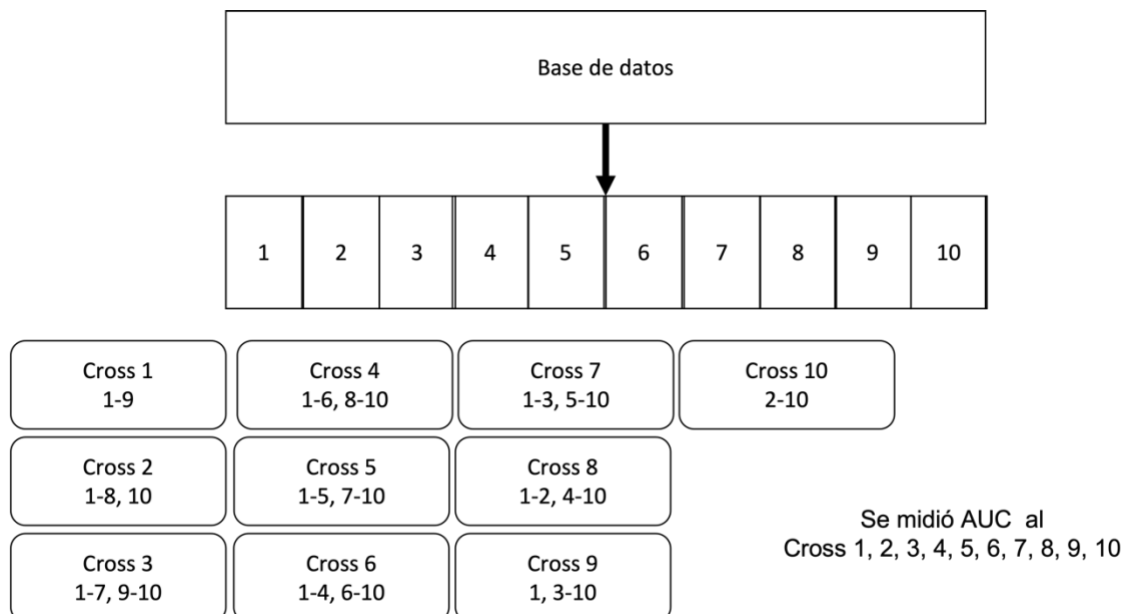
Los criterios utilizados para definir al mejor modelo predictor en la base de entrenamiento fueron: Menor número de variables, menor AIC y mayor AUC. Una vez identificado este modelo, se le realizó una pre-validación denominada “10-cross validation” o validación cruzada x 10 (Gerds TA, Kattan MW. Chapter 5. Does my model predict accurately? Medical Risk Prediction Models. With Ties to Machine Learning. 1st Edition. CRC Press. Pp 103-149 & Chapter 7. What would me make an expert? Medical Risk Prediction Models. With Ties to Machine Learning. 1st Edition. CRC Press. Pp 180-223). Para entender mejor esta validación cruzada, se describen a continuación los pasos:

Paso 1. Se divide la base de datos en 10 subgrupos generando números aleatorios del 1 al 10, para que cada subgrupo cuente con el mismo número de pacientes. Por ejemplo, si se tienen 100 pacientes, se realizan 10 subgrupos de 10 pacientes cada uno.

Paso 2. Se calcula el AUC dejando “fuera” uno de los 10 subgrupos. Se repite este paso dejando “fuera” cada uno de los 10 subgrupos. En este paso se midió el AUC en 10 ocasiones; en cada medición del AUC se dejó un subgrupo distinto fuera.

Paso 3. Si en esta validación cruzada se obtiene $AUC \geq 0.7$ en ≥ 9 de ellas, se considera que el desempeño del modelo predictor es adecuado o estable.

El siguiente esquema ejemplifica el proceso de la validación cruzada x 10:



X. RESULTADOS

El departamento de archivo clínico proporcionó 6,019 registros de expedientes clínicos electrónicos de pacientes con diagnóstico de cáncer de mama atendidos en el INCAN, de los cuales 5,481 fueron diagnosticados en el periodo de estudio, 5,009 cumplieron con los criterios de elegibilidad y 472 no.

Los motivos por los cuales los pacientes no fueron incluidos fueron: seguimiento ≤ 1 mes (n = 116), ausencia de etapa clínica o reporte de patología (n = 86).

Las razones de exclusión de pacientes fueron: cáncer de mama bilateral (n = 134), diagnóstico de metástasis cerebrales en los primeros 3 meses desde el diagnóstico de cáncer de mama (n = 44), otra histología p.ej. LNH, Phyllodes, sarcoma (n = 43), presencia de un segundo tumor primario (n = 41) y enfermedad neurológica previa (n = 8).

Datos faltantes: Se identificaron datos faltantes en 87 expedientes (1.8%) para la determinación de Ki-67 y en 262 expedientes (5.2%) para el SBR. Estos valores fueron imputados con la mediana del valor de la variable. Para Ki-67, el valor imputado fue 20% y para SBR fue de 7 (SBR II).

Cohorte

La cohorte analizada incluyó 5,009 pacientes, todas las variables que se reportan fueron obtenidas al diagnóstico de cáncer de mama. En breve, las variables clínicas se presentaron de la siguiente manera: La mediana de edad fue de 51.7 años (RIC 44.0 - 60.8), la mediana de KPS fue de 90, el tamaño del tumor fue ≤ 2 cm en 27.4% (n 1370), $>2 - 5$ cm en 46.2% (n 2316) y >5 cm en 26.4% (1323), se determinó la presencia de adenopatía axilar en el 53.4% (n 2673) y la etapa clínica anatómica fue catalogada como I en el 12.1% (n 605), II en 47.8% (n 2396), III en el 26.9% (n1347) y IV en el 13.2% (n 661); finalmente, el 56.6% (n 2833) se encontraba al diagnóstico en una etapa postmenopáusica. Las variables patológicas se encontraron de la siguiente manera: El 73.5% (n 3683) tuvo sobreexpresión de receptores de estrógeno (ER), 70.8 (n 2546) sobreexpresión de receptores de progesterona (PR) y 20.7% (n 1035) sobreexpresión de HER2. La proteína Ki-67 fue <10 en el 18.7%, 10-19 en 25.8%, 20-40 en 32.6% y >40 en 22.9%. El SBR fue I en 20.4%, II en 44.9% y III en 34.6%.

Al momento de la última revisión de expedientes clínicos, 787 pacientes (15.7%) habían fallecido y se hizo el diagnóstico de metástasis cerebrales a 287 (5.7%). Las características de la corte se presentan en la tabla R1.

Tabla R1. Características al diagnóstico de cáncer de mama en 5,009 pacientes

Variable	Unidad de medición	Total n 5009	Metástasis cerebrales n 287
Edad	Mediana (RIC), años	51.7 (44.0 – 60.8)	48.8 (41.7 – 56.9)
KPS al diagnóstico	Mediana (RIC)	90 (90- 90)	90 (90 – 90)
Tamaño tumor	≤ 2 cm, n (%)	1370 (27.4)	29 (2.1%)
	> 2 - 5 cm, n (%)	2316 (46.2)	84 (3.6%)
	> 5 cm, n (%)	1323 (26.4)	174 (13.2%)
Adenopatía axilar	Presente, n (%)	2673 (53.4)	227 (8.5%)
	Ausente, n (%)	2336 (46.6)	60 (2.6%)
Etapa clínica	I, n (%)	605 (12.1)	12 (2%)
	II, n (%)	2396 (47.8)	56 (2.3%)
	III, n (%)	1347 (26.9)	113 (8.4%)
	IV, n (%)	661 (13.2)	106 (16%)
Menopausia	Pre-menopausia, n (%)	2176 (43.4)	147 (6.8%)
	Pos-menopausia, n (%)	2833 (56.6)	140 (4.9%)
ER	Positivo, n (%)	3683 (73.5)	149 (4%)
	Negativo, n (%)	1326 (26.5)	138 (10.4%)
PR	Positivo, n (%)	3546 (70.8)	180 (4.5%)
	Negativo, n (%)	1463 (29.2)	127 (8.7%)
HER2	Positivo, n (%)	1035 (20.7)	106 (10.2%)
	Negativo, n (%)	3974 (79.3)	181 (4.6%)
Ki-67	<10, n (%)	936 (18.7%)	14 (1.5%)
	10-19, n (%)	1292 (25.8%)	46 (3.6%)
	20-40, n (%)	1635 (32.6%)	133 (8.1%)
	>40, n (%)	1146 (22.9%)	94 (8.2%)
SBR	SBR I, n (%)	1023 (20.4)	39 (3.8%)
	SBR II, n (%)	2251 (44.9)	94 (4.2%)
	SBR III, n (%)	1735 (34.6)	154 (8.9%)
Muerte	No, n (%)	4222 (84.3)	22 (0.5%)
	Si, n (%)	787 (15.7)	265 (33.7%)
Metástasis cerebrales n (%)	Ausentes, n (%)	4722 (94.3)	-
	Presentes, n (%)	287 (5.7)	287 (100%)

División de la cohorte

Se realizó la división simple de la cohorte, tras generar números aleatorios 1-1-2. De la cohorte con 5,009 pacientes, se incluyeron en la base de entrenamiento (BaseE) 3,339 (66.6%) y 1670 (33.3%) en la base de validación (BaseV).

Base de entrenamiento

Se midió la asociación entre las variables y el desenlace (metástasis cerebrales) calculando una regresión logística simple. Se obtuvo como medidas de asociación la razón de momios (OR), su intervalo de confianza al 95% (95%CI), y el valor de la prueba estadística de Wald (P), los hallazgos se presentan en la tabla R2. Todas las variables estudiadas demostraron una asociación significativa con el diagnóstico de metástasis cerebrales a excepción de la etapa clínica anatómica II [OR 1.18 (IC95% 0.63 – 2.22) P = 0.602] y el SBR II [OR 1.1 (IC95% 0.75-1.6), P = 0.625].

Tabla R2. Asociaciones simples por regresión logística entre las variables estudiadas y el desenlace (metástasis cerebrales).

Variable	OR (95%CI)	P
Edad	0.97 (0.96 – 0.98)	<0.0001
Premenopausia	1.39 (1.09 – 1.76)	0.006
Posmenopausia	Reference	-
Tamaño tumoral		
≤ 2 cm	Reference	-
2-5 cm	1.74 (1.13 – 2.66)	0.011
> 5cm	7.00 (4.69 – 10.45)	<0.0001
Adenopatía axilar presente	3.52 (2.63 – 4.70)	<0.0001
Adenopatía axilar negativa	Reference	-
Etapa clínica anatómica		
I	Reference	-
II	1.18 (0.63 – 2.22)	0.602
III	4.52 (2.47 – 8.27)	<0.0001
IV	9.43 (5.13 – 17.33)	<0.0001
ER positivo	Reference	-
ER negativo	2.75 (2.16 – 3.50)	<0.0001
PR positivo	Reference	-
PR negativo	2.01 (1.58 – 2.56)	<0.0001
HER2 positivo	2.39 (1.86 – 3.07)	<0.0001
HER2 negativo	Reference	-
Ki-67		
<10	Reference	-
10-19	2.17 (1.01 – 4.65)	0.046
20-40	6.10 (3.05 – 12.20)	<0.0001
≥40	5.93 (2.93-12.05)	<0.0001
SBR		
I	Reference	-
II	1.10 (0.75 – 1.60)	0.625
III	2.45 (1.71 – 3.52)	<0.0001

Correlaciones:

Se creó una matriz de correlaciones entre las variables como se puede observar en la Tabla R3. Se encontró una correlación significativa entre las variables ER y PR, por lo que se decidió excluir a la variable PR en la construcción del modelo predictor.

Tabla 3. Correlaciones no paramétricas (coeficientes de correlación) entre las variables asociadas a metástasis cerebrales.

	Edad	ER	HER2	T ≤2	T >2-5	T >5	Axila	EC I	EC II	EC III	EC IV	Ki67 <10	Ki67 10-19	Ki67 20-40	Ki67 >40	SBR I	SBR II	SBR III	
Edad	1																		
ER	0.058	1																	
PR	0.033	0.720																	
HER2	0.057	0.210	1																
T≤2	0.031	0.120	0.072	1															
T2-5	0.041	0.014	0.041	0.580	1														
T>5	0.079	0.139	0.121	0.359	0.553	1													
Axila	0.040	0.114	0.110	0.249	0.034	0.294	1												
EC I	0.029	0.104	0.065	0.604	0.350	0.217	0.393	1											
EC II	0.006	0.087	0.102	0.086	0.429	0.403	0.325	0.361	1										
EC III	0.032	0.129	0.090	0.246	0.115	0.383	0.542	0.220	0.579	1									
EC IV	0.022	0.062	0.098	0.136	0.147	0.307	0.155	0.144	0.379	0.231	1								
Ki67 <10	0.094	0.237	0.193	0.157	0.011	0.147	0.148	0.160	0.070	0.124	0.095	1							
Ki67 10-19	0.043	0.258	0.107	0.078	0.008	0.088	0.058	0.053	0.069	0.081	0.048	0.283	1						
Ki67 20-40	0.026	0.024	0.189	0.093	0.012	0.080	0.084	0.088	0.050	0.064	0.076	0.334	0.410	1					
Ki67 >40	0.104	0.461	0.079	0.124	0.011	0.138	0.104	0.106	0.081	0.128	0.053	0.261	0.321	0.379	1				
SBR I	0.079	0.227	0.155	0.116	0.022	0.092	0.112	0.121	0.038	0.100	0.042	0.303	0.084	0.136	0.217	1			
SBR II	0.023	0.175	0.005	0.030	0.036	0.072	0.047	0.030	0.038	0.062	0.004	0.032	0.136	0.019	0.193	0.458	1		
SBR III	0.091	0.375	0.136	0.129	0.020	0.153	0.144	0.133	0.072	0.149	0.040	0.291	0.213	0.095	0.386	0.369	0.658	1	

Modelos predictores

Se crearon distintos modelos predictores con regresión logística múltiple. A cada modelo construido se le midió el desempeño con el criterio de información de Akaike (AIC), el área bajo la curva (AUC) y el estadístico C (P). Las variables incluidas en cada modelo fueron combinaciones de las que se encontraron asociadas al desenlace descritas en la Tabla R2 y que no tuvieran una correlación significativa entre ellas. Los modelos construidos que presentaron la menor AIC y mayor AUC se describen en las tablas R4.

Tablas R4. Medición del desempeño predictor de los modelos construidos a partir de las combinaciones de variables asociadas a metástasis cerebrales en el análisis de regresión logística simple.

Modelo	1	2	3	4	5	6
V A R I A B L E S	Edad continua	Edad continua	Edad continua	Edad <40 40-60 >60	Edad <40 40-60 >60	Edad ≥44 45-50 >50-60 >60
	ER – vs. + HER 2 + vs. -	ER – vs. +	ER - vs. +	ER – vs. +	ER – vs. +	ER – vs. +
	>5 vs. ≤ 5 cm	>5 vs. ≤ 5 cm	>5 vs. ≤ 5 cm	>5 vs. ≤ 5 cm	>5 vs. ≤ 5 cm	>5 vs. ≤ 5 cm
	Adenopatía axilar + vs. -	Adenopatía axilar + vs. -	Adenopatía axilar + vs. -	Adenopatía axilar + vs. -	Adenopatía axilar + vs. -	Adenopatía axilar + vs. -
	EC Local L. avanzada Metastásica	EC Local L. avanzada Metastásica	Metastásica vs. no	Metastásica vs. no	Metastásica vs. no	Metastásica vs. no
	Ki-67 <15 15-45 >45	Ki-67 <15 15-45 >45	Ki-67 <15 15-45 >45	Ki-67 <15 15-45 >45	Ki-67 continua	Ki-67 continua
	SBR III	SBR III	SBR III	SBR III	SBR III	SBR III
AUC	0.780 (0.746 - 0.813)	0.780 (0.747-0.813)	0.779 (0.746-0.813)	0.778 (0.744-0.811)	0.765 (0.730-0.801)	0.766 (0.731-0.802)
AIC	1259	1259	1259	1261	1272	1271

Modelo	7	8	9	10	11	12
V A R I A B L E S	Edad continua	Edad continua	Edad continua	Edad continua	Edad continua	Edad continua
	TT <2cm 2-5cm >5cm	TT <2cm 2-5cm >5cm	TT <2cm 2-5cm >5cm	TT >5cm	TT <2cm 2-5cm >5cm	TT <2cm 2-5cm >5cm
	Adenopatía axilar	Adenopatía axilar	Adenopatía axilar	Adenopatía axilar	Adenopatía axilar	Adenopatía axilar
	EC Local L. avanzada Metastásica	EC I II III IV	EC I II III IV	EC I II III IV	EC Local L. avanzada Metastásica	EC I II III IV
	ER HER2	ER	ER HER2	ER HER2	ER HER2	ER HER2
	Ki-67 continua	Ki-67 continua	Ki-67 continua	Ki-67 continua	Ki-67 continua	Ki-67 continua
	SBR I II III	SBR I II III	SBR III	SBR III	SBR III	SBR I II III
	AUC	0.774 (0.739-0.808)	0.775 (0.741-0.809)	0.774 (0.740-0.809)	0.773 (0.738-0.808)	0.772 (0.738-0.807)
AIC	1261	1261	1264	1265	1267	1258

Modelo	13	14	15	16	17	18
V A R I A B L E S	Edad continua	Edad continua	Edad continua	Edad continua	Edad continua	Edad >60 >50-60 45-50 <45
	TT <2cm 2-5cm >5cm	TT <2cm 2-5cm >5cm	TT <2cm 2-5cm >5cm	TT <2cm 2-5cm >5cm	TT <2cm 2-5cm >5cm	TT <2cm 2-5cm >5cm
	Adenopatía axilar	Adenopatía axilar	Adenopatía axilar	Adenopatía axilar	Adenopatía axilar	Adenopatía axilar
	EC I II III IV	EC I II III IV	EC I II III IV	EC I II III IV	EC I II III IV	EC I II III IV
	ER	ER HER2	ER HER2	ER HER2	ER	ER
	Ki-67 continua	Ki-67 <10 10-19 20-40 >40	Ki-67 <10 10-19 20-40 >40	Ki-67 <20 ≥ 20	Ki-67 <10 10-19 20-40 >40	Ki-67 <10 10-19 20-40 >40
	SBR I II III	SBR I II III	SBR III	SBR III	SBR I II III	SBR I II III
AUC	0.775 (0.741-0.809)	0.792 (0.760-0.824)	0.785 (0.752-0.818)	0.781 (0.747-0.814)	0.793 (0.761-0.825)	0.790 (0.758-0.823)
AIC	1261	1241	1252	1256	1241	1243

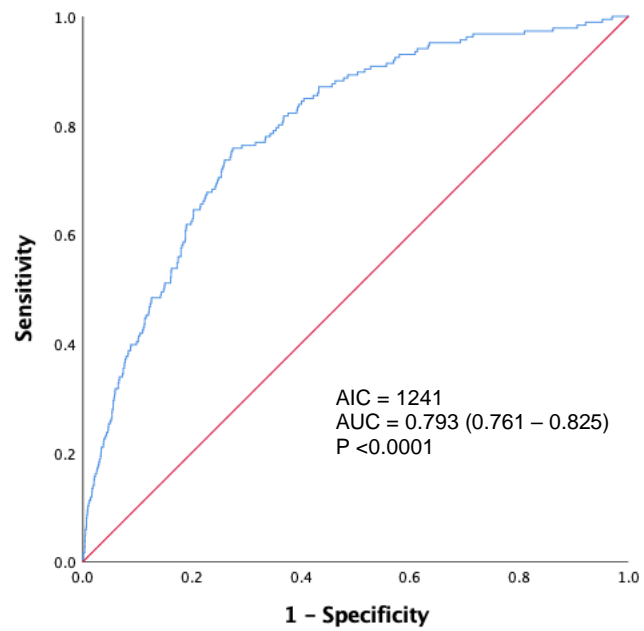
Modelo	19	20	21
V A R I A B L E S	Edad continua	Edad continua	Edad continua
	TT <2cm 2-5cm >5cm	TT <2cm 2-5cm >5cm	Pre/postmeno pausia
	Adenopatía axilar	Adenopatía axilar	TT <2cm 2-5cm >5cm
	EC Local L. avanzada Metastásica	EC Local L. avanzada Metastásica	Adenopatía axilar
	ER	ER	EC Local L. avanzada Metastásica
		HER2	ER
	Ki-67 <10 10-19 20-40 >40	Ki-67 <10 10-19 20-40 >40	Ki-67 <10 10-19 20-40 >40
SBR I II III	SBR I II III	SBR I II III	
AUC	0.791 (0.760-0.823)	0.791 (0.759-0.823)	0.794 (0.763-0.825)
AIC	1243	1243	1243

El modelo con menor AIC fue el construido a partir de las siguientes variables: Edad (variable continua), tamaño tumoral (≤ 2 cm, $>2 - 5$ cm, o > 5 cm), adenopatía axilar (presente o ausente), etapa clínica anatómica (I, II, III, o IV), receptores de estrógeno (ER positivo o negativo), Ki-67 (<10, 10-19, 20-40 o ≥ 40) y SBR (I, II o III). Los valores de β , las OR, su IC95% y valor de P se presentan en la Tabla R5 y la medición de su desempeño predictor en la figura R1.

Tabla R5. Modelo predictor de metástasis cerebrales que demostró lograr el mejor desempeño en la base de entrenamiento.

Variabes	β	OR (95%CI)	P
Edad	-0.015	0.98 (0.97 – 0.99)	0.017
Tamaño tumoral ≤2cm	-	Reference	-
2-5 cm	0.241	1.27 (0.69-2.33)	0.437
>5 cm	0.957	2.60 (1.40-4.81)	0.002
Adenopatía axilar Presente	-	Reference	-
Ausente	0.709	2.03 (1.25-3.28)	0.004
Etapa clínica I	-	Reference	-
II	-0.527	0.59 (0.22 - 1.52)	0.277
III	-0.161	0.85 (0.29 - 2.47)	0.767
IV	0.679	1.97 (0.69 – 5.57)	0.200
ER Positivo	-	Reference	-
Negativo	0.633	1.88 (1.31 – 2.69)	0.001
Ki-67 <10	-	Reference	-
10-19	0.645	1.90 (0.87 – 4.15)	0.105
20-40	1.30	3.66 (1.75 – 7.66)	0.001
>40	0.860	2.36 (1.06 – 5.22)	0.034
SBR I	-	Reference	-
II	-0.736	0.47 (0.29 – 0.76)	0.002
III	-0.346	0.70 (0.43 – 1.13)	0.155
Constante	-3.591	-	-

Figura R1. Curva de sensibilidad y 1-especificidad (ROC) para el desempeño del modelo predictor en la base de entrenamiento.



Validación cruzada

Se realizó la validación cruzada (10-fold cross-validation) del modelo predictor en la base de entrenamiento, midiendo el AUC para cada cruce tal y como se presenta en la tabla R6.

Tabla R6. Validación cruzada del modelo que presentó el mejor desempeño predictor

Cross	P	AUC	IC95%	
1 (n = 3005)	<0.0001	0.800	0.768	0.833
2 (n = 3005)	<0.0001	0.799	0.768	0.830
3 (n = 3005)	<0.0001	0.787	0.752	0.821
4 (n = 3005)	<0.0001	0.794	0.760	0.828
5 (n = 3005)	<0.0001	0.795	0.762	0.828
6 (n = 3005)	<0.0001	0.790	0.756	0.825
7 (n = 3005)	<0.0001	0.795	0.762	0.829
8 (n = 3005)	<0.0001	0.794	0.760	0.828
9 (n = 3005)	<0.0001	0.782	0.747	0.816
10 (n = 3006)	<0.0001	0.797	0.763	0.830
Promedio	<0.0001	0.793	0.759	0.826

Se concluye en esta etapa que el modelo predictor es estable y válido, ya que en todas las mediciones del AUC, se obtuvo un valor >0.70.

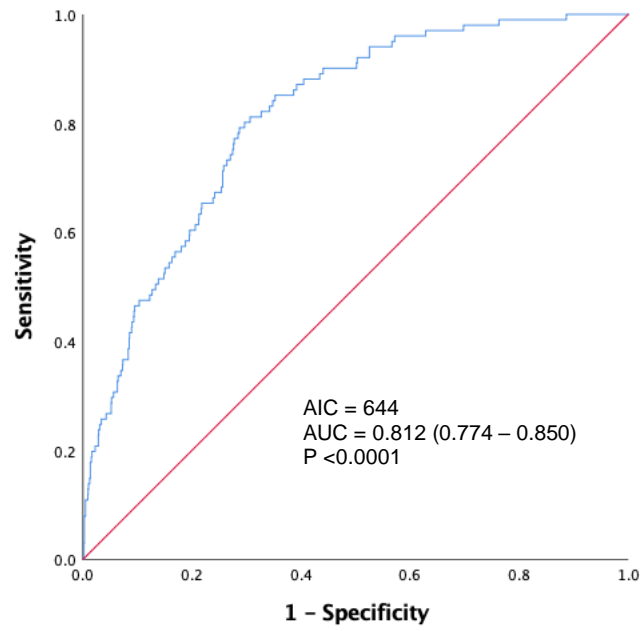
Desempeño del modelo predictor de metástasis cerebrales en la base de validación

Se midió el desempeño del modelo predictor en la base de validación (n = 1,670), los valores de la β para cada variable se presentan en la tabla R7 y el desempeño predictor en la Figura R2. Se concluye en esta etapa que el modelo presentó un adecuado desempeño medido por un AIC de 644, un AUC de 0.812 (IC95% 0.774 – 0.850), P <0.001.

Tabla R7. Regresión logística múltiple de las variables incluidas en el modelo predictor.

Variables	β	OR (95%CI)	P
Edad	-0.026	0.97 (0.95-0.99)	0.004
Tamaño tumoral			
≤2cm	-	Reference	-
2-5 cm	1.323	3.75 (1.11-12.68)	0.033
>5 cm	2.034	7.64 (2.29-25.43)	0.001
Adenopatía axilar			
Presente	-	Reference	-
Ausente	0.473	1.60 (0.84-3.03)	0.147
Etapa clínica			
I	-	Reference	-
II	-1.557	0.21 (0.42 - 1.04)	0.057
III	-1.157	0.31 (0.05 – 1.76)	0.188
IV	-0.084	0.91 (0.17 – 4.87)	0.921
ER			
Positivo	-	Reference	-
Negativo	0.707	2.02 (1.24 – 3.29)	0.004
Ki-67			
<10	-	Reference	-
10-19	0.580	1.78 (0.63 – 5.00)	0.270
20-40	0.484	1.62 (0.59 – 4.42)	0.345
>40	0.173	1.18 (0.41 – 3.44)	0.750
SBR			
I	-	Reference	-
II	0.802	2.22 (0.89 – 5.52)	0.084
III	1.118	3.05 (1.21 – 7.70)	0.018
Constante	-3.767	-	-

Figura R2. Desempeño predictor del modelo predictor en la base de validación



Interpretación clínica del modelo

Para que el modelo pueda emplearse en el ejercicio clínico se hicieron los siguientes pasos:

- Obtención de los valores β del modelo predictor en la cohorte,
- Cálculo de la posibilidad individual de metástasis cerebrales, y
- División en grupos de acorde a la posibilidad de metástasis cerebrales.

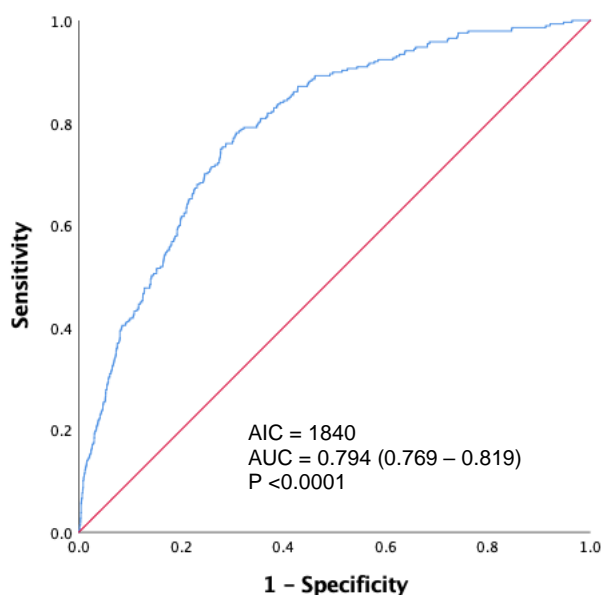
a) Obtención de los valores β del modelo en toda la cohorte:

Se hizo una regresión logística múltiple con las variables predictoras para obtener el valor de b y la constante como se presenta en la tabla R8. Además, se midió el desempeño del modelo en la cohorte, obteniendo un AIC de 1840, un AUC de 0.794 (IC95% 0.769 – 0.819), $P < 0.0001$, como se presenta en la figura R3.

Tabla R8. Regresión logística múltiple para obtener los valores β en la cohorte.

Variables	β	OR (95%CI)	P
Edad	-0.018	0.98 (0.97-0.99)	<0.0001
Tamaño tumoral			
≤2cm	-	Reference	-
2-5 cm	0.539	1.71 (1.002-2.93)	0.049
>5 cm	1.251	3.49 (2.04-5.97)	<0.0001
Adenopatía axilar			
Ausente	-	Reference	-
Presente	0.623	1.86 (1.27-2.73)	0.001
Etapas clínicas			
I	-	Reference	-
II	-0.812	0.44 (0.19 – 0.99)	0.048
III	-0.427	0.65 (0.26 – 1.59)	0.348
IV	0.467	1.59 (0.67 – 2.78)	0.290
ER			
Positivo	-	Reference	-
Negativo	0.640	1.89 (1.42 – 2.52)	<0.0001
Ki-67			
<10	-	Reference	-
10-19	0.633	1.88 (1.01 – 3.50)	0.045
20-40	1.008	2.74 (1.51 – 4.95)	0.001
>40	0.625	1.86 (0.98 – 3.52)	0.054
SBR			
I	-	Reference	-
II	-0.317	0.72 (0.48 – 1.09)	0.124
III	0.044	1.04 (0.69 – 1.57)	0.834
Constante	-3.540	-	-

Figura R3. Desempeño del modelo predictor en la cohorte



b) Cálculo de la posibilidad individual de metástasis cerebrales.

Se utilizó la fórmula para el cálculo de la posibilidad de metástasis cerebrales a cada paciente sustituyendo las variables:

$$P = \text{Constante} + \beta_1(X_1) + \beta_2(X_2) + \beta_3(X_3) + \beta_4(X_4) + \beta_5(X_4) + \beta_6(X_6) + \beta_7(X_7)$$

$$P = -3.54 + \text{Edad} + \text{TT} + \text{Adenopatía axilar} + \text{EC} + \text{ER} + \text{Ki-67} + \text{SBR}$$

Variable	Unidad	Valor para la ecuación
Edad	Edad en años (dos decimales)	Edad *-0.018
Tamaño tumoral (TT)	≤ 2 cm	0
	>2cm – 5 cm	0.539
	>5cm	1.251
Adenopatía axilar	Ausente	0
	Presente	0.623
Etapa clínica anatómica (EC)	I	0
	II	-0.812
	III	-0.427
	IV	0.467
Receptores de estrógeno (ER)	Positivos	0
	Negativos	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

c) División en grupos de acorde a la posibilidad individual de metástasis cerebrales

El resultado de la ecuación mencionada (posibilidad individual) se dividió en deciles, cuartiles y terciles. La incidencia relativa de acorde a cada uno de los valores de corte se presenta en las Tablas R9, R10 y R11.

Tabla R9. Incidencia relativa de metástasis cerebrales en la cohorte, de acuerdo al resultado de la ecuación obtenida por el modelo predictor, dividido en deciles.

Deciles	Valor de la ecuación	n	Metástasis cerebrales, n(%)
0	≤ -4.7242	500	4 (0.8%)
1	-4.7243 a -4.3401	502	2 (0.4%)
2	-4.3402 a -4.0378	500	7 (1.4%)
3	-4.0379 a -3.7687	502	10 (2.0%)
4	-3.7688 a -3.4605	500	8 (1.6%)
5	-3.4605 a -3.1411	501	21 (4.2%)
6	-3.1412 a -2.7559	502	27 (5.4%)
7	-2.7560 a -2.3175	501	48 (9.6%)
8	-2.3175 a -1.7321	501	47 (9.4%)
9	≥ -1.7322	500	113 (22.6%)

Tabla R10. Incidencia relativa de metástasis cerebrales en la cohorte, de acuerdo al resultado de la ecuación obtenida por el modelo predictor, dividido en cuartiles.

CUARTILES	Valor de la ecuación	n	Metástasis cerebrales, n(%)
0	≤ -4.1944	1252	10 (0.8%)
1	-4.1945 a -3.4605	1252	21 (1.7%)
2	-3.4606 a -2.5317	1253	66 (5.3%)
3	≥ -2.5318	1252	190 (15.2%)

Tabla R11. Incidencia relativa de metástasis cerebrales en la cohorte, de acuerdo al resultado de la ecuación obtenida por el modelo predictor, dividido en terciles.

TERCILES	Valor de la ecuación	n	Metástasis cerebrales, n(%)
0	≤ -3.9564	1653	17 (1.0%)
1	-3.9565 A -2.9188	1653	46 (2.8%)
2	≥ -2.9189	1703	224 (13.2%)

Ante estos hallazgos, se decidió dividir en 4 grupos de acorde a la posibilidad individual de metástasis cerebrales determinado por el resultado de la ecuación. Los valores de la ecuación, la incidencia acumulada y la significancia estadística se presentan en la tabla R12.

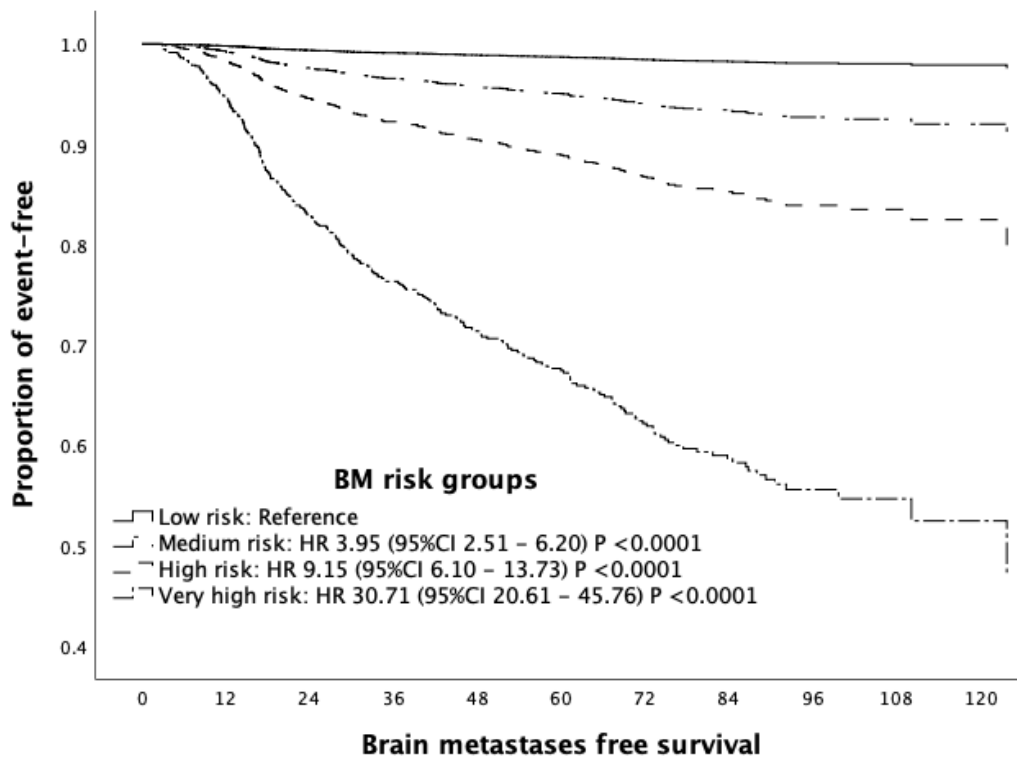
Tabla R12. Valor de la ecuación, incidencia acumulada y significancia estadística para cada grupo acorde a la posibilidad de diagnóstico de metástasis cerebrales.

Grupos acorde a la posibilidad de metástasis cerebrales	Valor de la ecuación	n	Metástasis Cerebrales n (%)	Wald test P
Baja	≤ -3.4605	2504	31 (1.2%)	Reference
Moderada	-3.4604 to -2.7559	999	48 (4.8%)	<0.0001
Elevada	-2.7558 to -1.7321	1002	95 (9.5%)	<0.0001
Muy elevada	≥ -1.7320	500	113 (22.6%)	<0.0001

INTERPRETACIÓN DEL MODELO

En la figura R4 se puede observar las diferencias entre los grupos de acorde a su posibilidad de tener metástasis cerebrales; el grupo de bajo riesgo (low risk), riesgo medio (medium risk), riesgo alto (high risk) y riesgo muy alto (very high risk) de tener el diagnóstico de metástasis cerebrales.

Figura R4. Curva de regresión de Cox para medir el riesgo de ser diagnosticada con metástasis cerebrales de acorde a los grupos de riesgo.



Explicación de la categorización de las variables.

Edad

La variable edad no tuvo una distribución normal ($K-S < 0.05$). La mediana de edad al diagnóstico de cáncer de mama fue de 51.75 años, rango intercuartil 44.03-60.82 años, edad mínima de 20.30 años y máxima de 93.87 años. Se dividió la edad la edad en deciles, cuartiles y terciles para conocer la prevalencia de metástasis cerebrales para cada grupo como se demuestra en la Tabla RR 12,13 y 14.

Tabla RR12. Incidencia acumulada de metástasis cerebrales en la cohorte de acorde a la edad al diagnóstico de cáncer de mama.

Decil	Rango de edad en años	n	Metástasis cerebrales n (%)
0	≤ 37.19	545	46 (8.4%)
1	37.20-42.14	456	30 (6.6%)
2	42.15-45.65	500	34 (6.8%)
3	45.66-48.52	502	30 (6.0%)
4	48.53-51.75	502	33 (6.6%)
5	51.76-54.97	499	30 (6.0%)
6	54.98-58.62	503	24 (4.8%)
7	58.63-63.15	502	27 (5.4%)
8	63.16-70.18	499	19 (3.8%)
9	≥70.19	501	14 (2.8%)

Tabla RR13. Incidencia acumulada de metástasis cerebrales en la cohorte de acorde a la edad al diagnóstico de cáncer de mama.

Cuartil	Rango de edad en años	n	Metástasis cerebrales n (%)
0	≤44.03	1252	96 (7.7%)
1	44.04 - 51.75	1253	77 (6.1%)
2	51.76 - 60.82	1252	65 (5.2%)
3	≥60.83	1252	49 (3.9%)

Tabla RR14. Incidencia acumulada de metástasis cerebrales en la cohorte de acorde a la edad al diagnóstico de cáncer de mama.

Tercil	Rango de edad en años	n	Metástasis cerebrales n (%)
0	≤46.46	1653	123 (7.4%)
1	46.47 – 57.14	1653	95 (5.7%)
2	≥57.15	1703	69 (4.1%)

Se ingresó la variable edad como variable continua, en deciles, en cuartiles y terciles al modelo predictor. Se encontró que el desempeño del modelo fue mejor cuando se ingresó como variable continua tal y como se aprecia en la Tabla RR 15.

Tabla R15. Medición del desempeño del modelo predictor con la variable edad ingresada de manera continua, en deciles, cuartile y terciles.

	AIC	AUC (IC95%)	P
Variable continua	1840	0.794 (0.769 – 0.819)	<0.0001
Deciles	1885	0.794 (0.769-0.819)	<0.0001
Cuartiles	1892	0.792 (0.767-0.818)	<0.0001
Terciles	1893	0.792 (0.766 – 0.817)	<0.0001

Otra razón por la cual se mantuvo esta variable como continua al construir el modelo fue que presentó una relación lineal con el desenlace medido con regresión lineal (R^2 0.004, $P < 0.0001$).

Ki-67

La variable Ki-67 no tuvo una distribución normal (K-S <0.05), la mediana de su valor fue de 20%, el rango intercuartil 10% – 40%, valor mínimo 1% y máximo de 95%, la percentil 33 fue de 10% y la percentil 66 de 30%. Se dividió el valor de Ki-67 en deciles, cuartiles, terciles y en su valor clínico de corte (<20 vs ≥20) para conocer la incidencia acumulada metástasis cerebrales en cada grupo como se demuestra en las Tablas RR 16-19.

Tabla RR16. Incidencia acumulada de metástasis cerebrales en la cohorte de acorde a Ki-67 en deciles al diagnóstico de cáncer de mama.

Decil	Rango del valor de Ki-67	n	Metástasis cerebrales n (%)
0	0-10	1796	38 (2.1%)
1	11-20	1139	74 (6.5%)
2	21-30	580	47 (8.1%)
3	31-40	348	34 (9.8%)
4	41-50	245	27 (11%)
5	51-60	316	15 (4.7%)
6	61-70	277	27 (9.7%)
7	71-80	208	17 (8.2%)
8	81-90	95	8 (8.4%)
9	≥91	5	0

Tabla RR17. Incidencia acumulada de metástasis cerebrales en la cohorte de acorde a Ki-67 en cuartiles al diagnóstico de cáncer de mama.

Cuartil	Rango del valor de Ki-67	n	Metástasis cerebrales n (%)
0	<10	936	14 (1.5%)
1	10-19	1292	46 (3.6%)
2	20-40	1635	133 (8.1%)
3	>40	1146	94 (8.2%)

Tabla RR18. Incidencia acumulada de metástasis cerebrales en la cohorte de acorde a Ki-67 en terciles al diagnóstico de cáncer de mama.

Tercil	Rango del valor de Ki-67	n	Metástasis cerebrales n (%)
0	0-10	1796	38 (2.1%)
1	11-30	1719	121 (7.0%)
2	>30	1494	128 (8.6%)

Tabla RR19. Incidencia acumulada de metástasis cerebrales en la cohorte de acorde a Ki-67 dividido entre el nivel de corte clínico (<20 vs ≥20).

Rango del valor de Ki-67	n	Metástasis cerebrales n (%)
<20	2228	60 (3.7%)
≥20	2781	227 (8.2%)

No se ha establecido un nivel de corte universal para el valor de Ki-67 como predictor de metástasis cerebrales. Se ingresó al modelo esta variable como continua, dividida en deciles, cuartiles, terciles y como <20 vs. ≥20. Se encontró que el desempeño del modelo fue mejor cuando se ingresó Ki-67 como variable dividida en cuartiles; por lo tanto, se decidió ingresar esta variable al modelo así siguiendo las recomendaciones de la literatura (Gerds TA, Kattan MW. Chapter 4. I am ready to build a prediction model. Medical Risk Prediction Models. With Ties to Machine Learning. 1st Edition. CRC Press. Pp 62-99). La comparación de los modelos predictores con distintos valores para Ki-67 se presentan en la Tabla R 20.

Tabla R20. Medición del desempeño del modelo predictor de metástasis cerebrales con la variable Ki-67 ingresada de manera continua, en deciles, cuartiles y terciles.

	AIC	AUC (IC95%)	P
Variable continua	1906	0.784 (0.757 – 0.810)	<0.0001
Deciles	1877	0.798 (0.771 – 0.823)	<0.0001
Cuartiles	1840	0.794 (0.769 – 0.819)	<0.0001
Terciles	1893	0.791 (0.765 – 0.817)	<0.0001
<20 vs ≥ 20	1900	0.788 /0.762 – 0.814)	<0.0001

XI. DISCUSIÓN

El cáncer de mama (CaMa) es la neoplasia maligna más frecuente a nivel mundial, en nuestra población (México), los resultados de GLOBOCAN 2020 reportan que el CaMa es la neoplasia con mayor número de casos nuevos por año con 29,929 (15% del total de neoplasias considerando ambos géneros y 28% en mujeres). El CaMa es la principal causa de muerte asociada a cáncer con 7,931(8.8%) de las muertes reportadas en el 2020. Todas las pacientes incluidas en la cohorte recibieron el tratamiento inicial recomendado por las guías NCCN y las recomendaciones descritas en la literatura(69).

La incidencia de metástasis cerebrales (MeCe) en pacientes con CaMa reportada en la literatura varía entre 0.35 – 24% en estudios retrospectivos, con una mediana de supervivencia que oscila entre 4-13 meses(70). La incidencia acumulada en nuestra cohorte fue de 287 pacientes con MeCe en 5009 pacientes con CaMa; es decir, se diagnosticó MeCe al 5.7% de los pacientes incluidos en la cohorte, similar a lo encontrado por otros reportes. La prevalencia de MeCe en distintos estudios publicados se resume en la tabla D1.

La mediana del periodo desde el diagnóstico de CaMa a MeCe (Supervivencia libre de metástasis cerebrales = SLMeCe) en la presente cohorte fue de 23.29 meses (IC95% 19.79 – 26.79) similar a lo reportado en otros estudios como se puede ver en la tabla D1.

Tabla D1. Resumen de la prevalencia y supervivencia libre de metástasis cerebrales en mujeres con cáncer de mama reportados en la literatura.

Tamaño de muestra	MeCe	SLMeCe	Referencia
9,524 CaMa temprano	517 (5.4%)	NR	Pestalozzi et al 2006 (71)
51,898	2,635 (5.1%)	NR	Barnholtz-Sloan et al 2004(72)
2193 CaMa	160 (7.3%)	NR	Azim et al 2018(39)
488 CMM	115 (23.6%)	31 meses (Rango 0 – 192)	Aversa et al 2014 (25)
10,782 CaMa	383 (3.55%)	NR	Tham et al 2006(73)
2,441 CaMa	80 (3.27%)	NR	Heitz et al 2009(74)
2,136 CMM	362 (16.94%)	8.9 meses (Rango 1-98)	Graesslin et al 2010(75)
626 CMM	66 (10.5%)	NR	Heitz et al 2011(76)
2248 CaMa	164 (7.3%)	35.3 meses (RIC 57.1 – 19.7)	Hung et al 2014(77)
238,726 CaMa	968 (0.41%)	NR	Martin et al 2017(18)
206,913 CaMa	738 (0.35%)	NR	Kim et al 2018(23)
209 MeCe CaMa	209	37 meses (Rango 0-432)	Brogi et al 2011(78)
5604 CaMa	18 (0.32%)	1.9 años	Shao et al 2011(79)
140 MeCe CaMa	140	46 meses (Rango 0 – 266)	Shen et al 2015(80)
63 MeCe CaMa	66	30.2 meses (Rango 21.5 – 38.8)	Dayan et al 2012 (81)
201 CaMa HER2+	27 (9%)	NR	Gabos et al 2006(82)
1,434 CaMa EC I/II	36 (2.5%)	51.4 meses (Rango 7.6 – 108)	Arvold et al 2012(26)
912 CaMa	30 (3.28%)	29 meses	Chow et al 2015(83)
591 CaMa ECI-III	17 (2.87%)	NR	Ishihara et al 2013(84)
206 CaMa	15 (7.3%)	NR	Xue et al 2013(85)
492 CMM	116 (23.57%) En realidad 70/446 (15.69%)	17 meses (Rango 2-107.6)	Genre et al 2017(46)
473 CMM	74 (15.6%)	40.3 meses (Rango 0 – 252)	Minisini et al 2013(86)
5,009 CaMa	287 (5.7%)	29.29 meses (IC95% 19.79 – 26.79) Rango 3-124 meses	Presente estudio

CaMa = Cáncer de mama, CMM = Cáncer de Mama Metastásico, EC = Etapa clínica, NR = No reportado, RIQ = rango intercuartil, SLMeCe = Supervivencia libre de metástasis cerebrales.

XI.A VARIABLES PREDICTORAS

Las variables asociadas a un mayor riesgo de diagnóstico de metástasis cerebrales en nuestra cohorte utilizadas para la construcción del modelo predictor de metástasis cerebrales fueron: Edad (variable continua), tamaño del tumor primario (TT ≤ 2 cm, $>2 - 5$ cm o >5 cm), adenopatía axilar (presente o ausente, etapa clínica anatómica (I, II, III, IV), expresión del receptor de estrógeno (positivo o negativo), porcentaje de expresión de Ki-67 (<10 , 10-19, 20-40 y $>40\%$) y SBR (I, II, III); todas estas variables han sido consistentemente asociadas a MeCe en la revisión de la literatura como se presenta a continuación.

XI.A1 Edad

El efecto de la edad en el riesgo de MeCe ha sido estudiado por varios autores como variable continua o categorizada. A pesar de la heterogeneidad en su medición y puntos de corte, una menor edad al diagnóstico de CaMa o CMM se ha asociado con MeCe, tanto en la asociación simple como en la múltiple. Es importante notar que Aversa et al. calculó un decremento en el riesgo de MeCe de 6% por cada año adicional a la edad al diagnóstico de CaMa(25). El presente estudio demostró una asociación entre la edad y el diagnóstico de MeCe tras regresión logística simple y la múltiple en la base de entrenamiento y la base de validación similar a la previamente reportada por otros estudios tal y como se describe en la Tabla D2.

Tabla D2. Resumen de estudios publicados que han determinado una asociación significativa entre la edad con MeCe en pacientes con cáncer de mama.

Edad	Medida de Asociación	Referencia
Continua (AM)	aHR= 0.97 (IC95% 0.95-0.99) P = 0.024	Arvold et al 2012 (26)
Categorizada (AM)	≤ 35 Referencia >35 OR 0.884 (IC95% 0.786 – 0.994) P = 0.02	Xue et al 2013(85)
Continua (AS)	HR = 0.94 p <0.0001	Aversa et al 2014(25)
Categorizada (AS)	<35 años 10.8% (40/372), 35-49 años 6.4% (225/3542), 50-59 años 5.6% (164/2936), ≥ 60 años 3.3% (88/2674) P <0.01	Pestalozzi et al 2006(71)
Continua (AM)	OR 0.98 (IC95% 0.97 – 0.99) P <0.001	Tham et al 2006(73)
Categorizada (AS)	<50 55% (30/55) vs >50 años 45% (25/55) P = 0.001	Hicks et al 2006(50)
Categorizada (AM)	< 50 años HR 2.0 (IC95% 1.2 – 3.5) P = 0.012	Heitz et al 2009(74)
Continua (AM)	OR 0.98 (IC95% 0.97 – 0.99) P <0.001	Graesslin et al 2010(75)
Categorizada (AM)	<50 años HR 1.97 (IC95% 1.18 – 3.30) P = 0.009	Heitz et al 2011(76)
Continua (AS)	Mediana MeCe 52 vs no MeCe 60 años P <0.001	Sosinska-Mielcarek et al 2013(87)
Categorizada (AM)	≤ 35 referencia vs >35 OR 0.884 (IC95% 0.786-0.994) P = 0.02	Xue et al 2013(85)
Categorizada (AM)	≥ 60 años Referencia ≤ 35 años HR 2.09 (IC95% 1.15 – 3.81) P = 0.016	Hung et al 2014(77)
Categorizada (AS)	Cohorte Total: 18-40 años 57/12,564 (0.45%) 41-60 años 423/98,184 (0.43%), 61-80 años 425/103,700 (0.40%), >80 años 63/23,265 (0.27%) CMM: 18-40 Ref, 41-60 OR 1.41 (IC95% 1.05 – 1.88) P = 0.02, 61-80 OR 1.40 (IC95% 1.04 – 1.88) P = 0.03, >80 NR	Martin et al 2017(88)
Categorizada (AS)	<40 vs ≥ 40 años P 0.04	Shen et al 2015(80)
Continua (AM)	OR 0.995 (IC95% 0.988 – 1.003) P 0.216	Kim et al 2018 (23)
Continua (AS)	Mediana MeCe 53 vs 56 P = 0.01 AM negativo	Genre et al 2017 (46)
Continua (AS)	OR 0.97 (IC95% 0.96 – 0.98) P < 0.0001	Base de entrenamiento
Continua (AM)	OR 0.98 (IC95% 0.97 – 0.99) P = 0.017	Base de entrenamiento
Continua (AM)	OR 0.97 (IC95% 0.95 – 0.99) P = 0.004	Base de validación
Continua (AM)	OR 0.98 (IC95% 0.97 – 0.99) P < 0.0001	Cohorte

*AM = análisis múltiple. *AS = análisis simple

XI.A2 Tamaño tumoral

El tamaño del tumor en la mama al diagnóstico de CaMa ha sido asociado al pronóstico ya que las pacientes con mayor tamaño tumoral (TT) tienen un peor pronóstico y viceversa. La AJCC ha categorizado el tamaño tumoral, como T1 para aquellas lesiones cuyo diámetro mayor sea ≤ 2 cm, T2 tumores de >2 pero ≤ 5 cm y T3 para los tumores >5 cm. La asociación del TT con el diagnóstico de MeCe del presente estudio y los principales estudios en la literatura médica se describen en la tabla D3.

Tabla D3. Resumen de estudios publicados que han determinado una asociación significativa entre el tamaño tumoral al diagnóstico de CaMa con MeCe.

Análisis	Medida Asociación	Referencia
AS	≤ 2 cm 3.9% (179/4545) >2 cm 6.9% (324/4678)	Pestalozzi et al 2006(71)
AM	≤ 2 cm Referencia 2.1-5 cm OR 1.5 (IC95% 1.1 – 2.0) >5.1 cm OR 1 (IC95% 0.7 – 1.5) P = 0.01	Tham et al 2006(73)
AM	pT3/4 HR 1.9 (IC95% 1.2 – 3.6) P = 0.02	Heitz et al 2009 (74)
AS	Mediana de SLMeCe (P = 0.03) T1 41.3 meses T2 26.7 meses T3 29.7 mees	Brogi et al 2011(78)
AS	T1 vs T2/T3 P = 0.002	Shen et al 2015(80)
AM	>2 cm HR 2.76 (IC95% 1.23 – 6.19) P 0.013 Pacientes HER2+	Gabos et al 2006 (82)
AS	pT (cm) OR 0.93 (IC95% 0.87 – 0.99) P = 0.03 AM negativa	Graesslin et al 2010(45)
AS	MeCe T1/2 48.5% vs T3/4 57.5% P =0.015 AM negativa	Heitz et al 2011(76)
AS	T0,1,2,3,4 P = 0.09 AM negativo	Genre et al 2017(46)
AM	>2 cm HR 3.60 (I95% 1.54 – 8.38) P = 0.003	Azim et al 2018(39)
AM	T ≤ 1 Referencia T2 OR 1.513 (IC95% 1.008 – 2.269) P = 0.045 T3 OR 2.167 (IC95% 1.366 – 3.436) P = 0.001 T4 OR 3.045 (IC95% 2.004 – 4.626) P <0.0001	Kim et al 2018(23)
AS	≤ 2 cm Referencia $> 2 - 5$ cm OR 1.74 (IC95% 1.13 – 2.66) P = 0.011 > 5 cm OR 7.00 (IC95% 4.69 – 10.45) P <0.0001	Base de entrenamiento
AM	≤ 2 cm Referencia $> 2 - 5$ cm OR 1.27 (IC95% 0.69 – 2.33) P = 0.437 > 5 cm OR 2.60 (IC95% 1.4 – 4.81) P = 0.002	Base de entrenamiento
AM	≤ 2 cm Referencia $> 2 - 5$ cm OR 3.75 (IC95% 1.11 – 12.68) P = 0.033 > 5 cm OR 7.64 (IC95% 2.29 – 25.43) P = 0.001	Base de validación

AM = Análisis múltiple, AS = análisis simple, cm = centímetros, OR = Razón de momios, IC = intervalo de confianza, pT = tamaño medido por estudio patológico.

XI.A3 Adenopatía axilar

La presencia de enfermedad ganglionar axilar (adenopatía axilar) se asocia con un peor pronóstico en pacientes con CaMa, al buscar intencionadamente en los expedientes electrónicos, fue difícil caracterizar el número y tamaño de ganglios linfáticos regionales afectados; por tal motivo, se decidió categorizar como presencia o ausencia de adenopatía axilar, tal y como lo han hecho otros autores.

En la Tabla D5 se describe la medición de la asociación entre la presencia de adenopatía axilar y metástasis cerebrales encontrados en este estudio y en la literatura médica.

Tabla D4. Resumen de estudios publicados que han determinado una asociación significativa entre la presencia de afección ganglionar con el diagnóstico de metástasis cerebrales en pacientes con cáncer de mama.

Análisis	Medidas de asociación	Referencia
AS	Ganglios negativos 18.8% (754/4007) 1-3 ganglios 41.9% (1405/3335) ≥4 ganglios 63.7% (1377/2163) P <0.01	Pestalozzi et al 2006(71)
AS	Adenopatía 71% (23/31) P = 0.001	Hicks et al 2006(50)
AM	Ganglios positivos HR 2.4 (IC95% 1.1 – 5.1) P = 0.02	Heitz et a 2009(74)
AS	Mediana SLMeCe (P = 0.02) Ganglios negativos 42.5 meses Ganglios positivos 33.8 meses	Brogi et al 2011(78)
AS	Ganglios positivos HR 2.84 (IC95% 1.40 – 5.78) P = 0.019 AM sin asociación	Gabos et al 2006(82)
AS	OR 0.76 (IC95% 0.59 – 0.98) P = 0.03 AM negativa	Graesslin et al 2010(45)
AM	>4 vs ≤3ganglios axilares OR 8.390 (IC95%1.314 – 53.558) P >0.001	Xue et al 2013(85)
AS	N0/1/2/3 P = 0.026 AM negativa	Chow et al 2015(83)
AM	N 0 Referencia N1 OR 1.085 (IC95% 0.808-1.458) P = 0.587 N2 OR 1.211 (IC95% 0.804 – 1.823) P =0.360 N3 OR 1.564 (IC95% 1.082 – 2.261) P = 0.017	Kim et al 2018(23)
AS	Ganglios Positivos 14/18 P <0.001	Shao et al 2011(79)
AM	HR 4.03 (IC95% 1.90-8.52) P = <0.001	Azim et al 2018(39)
AS	Adenopatía axillar + OR 3.52 (IC95% 2.63 – 4.70) P <0.0001	Base de entrenamiento
AM	Adenopatía axillar + OR 2.03 (IC95% 1.25 – 3.28) P = 0.004	Base de entrenamiento
AM	Adenopatía axillar + OR 1.60 (IC95% 0.84 – 3.03) P = 0.147	Base de validación

AM = Análisis múltiple, AS = Análisis simple (Univariable), CMM = Cáncer de mama metastásico, EC = Etapa clínica, HR = Hazard ratio, IC95% = Intervalo de confianza al 95% OR = Razon de Momios SLMeCe = supervivencia libre de metástasis cerebrales

XI.A4 Etapa clínica

La etapa clínica (EC) anatómica es una de las herramientas principales para guiar el tratamiento y la principal orientadora de pronóstico en mujeres con CaMa, a mayor EC se considera peor el pronóstico y requiere un tratamiento más intensivo. Es importante hacer notar que aunque no todas las EC fueron estadísticamente significativas, al incluir esta variable, el rendimiento del modelo predictor fue mejor. En la Tabla D5 se resume la medición de la asociación entre la EC y metástasis cerebrales encontrados en el presente estudio y en la literatura médica.

Tabla D5. Resumen de los estudios que han reportado asociación entre la etapa clínica (EC) y el diagnóstico de MeCe

Análisis	Medida de asociación	Referencia
AM	EC IV HR 2.64 (IC95% 1.31 – 5.33) P 0.007	Heitz et al 2011 (76)
AM	ECIV vs I HR 67.07 (IC95% 34.05 – 132.14) P <0.001	Hung et al 2014(77)
AS	EC I/II vs III/IV P <0.0001	Shen et al 2015(89)
AM	CMM aHR 28.46 (IC95% 12.3 – 65.7) P <0.001	Arvold et al 2012(26)
AS	Localizada 688/27,598 (2.5%) Regional 1,266/18,484 (6.8%) Distante 532/3,750 (14.2%)	Barnholtz-Sloan et al 2004(72)
AS	ECI Referencia EC II OR 1.18 (IC95% 0.63 – 2.22) P = 0.602 EC III OR 4.52 (IC95% 2.47 – 8.27) P <0.0001 EC IV OR 9.43 (IC95% 5.13 – 17.33) P <0.0001	Base de entrenamiento
AM	ECI Referencia EC II OR 0.59 (IC95% 0.22 – 1.52) P = 0.277 EC III OR 0.85 (IC95% 0.29 – 2.47) P = 0.767 EC IV OR 1.97 (IC95% 0.69 – 5.57) P = 0.200	Base de entrenamiento
AM	ECI Referencia EC II OR 0.21 (IC95% 0.42 – 1.04) P = 0.057 EC III OR 0.31 (IC95% 0.05 – 1.76) P = 0.188 EC IV OR 0.91 (IC95% 0.17 – 4.87) P = 0.921	Base de validación

AM = Análisis múltiple, AS = Análisis simple, CMM = Cáncer de mama metastásico, EC = Etapa clínica, HR = Hazzard ratio, IC95% = Intervalo de confianza al 95% OR = Razon de Momios

Estudios como el reportado por Martínez-Aranda et al 2017(90) solo mencionan que la edad <50 años (P = 0.001), RH- (P < 0.0001) y afección ganglionar (P < 0.0001) fueron más frecuentes en aquellos con MeCe, pero al no proporcionar la descripción de las medidas de asociación, no se describen en la tabla correspondiente a Edad, RE o afección ganglionar. Por otro lado, algunos estudios reportan la asociación simple significativa entre alguna o varias de las variables incluidas en el modelo (Edad, RE, Tamaño tumoral, adenopatía axilar y/o etapa clínica) y MeCe(23,26,34,45,50,79–81,84,86,87,91,92) pero no encontraron asociación múltiple significativa o no la reportan. Por ejemplo, Azim et al 2018(39) encontró asociación entre tamaño (P = 0.004), etapa ganglionar (P<0.001) y etapa clínica (P = 0.002) en el análisis simple pero no reportan sus resultados en el análisis múltiple; por esta razón, solo se comentan en las tablas correspondientes los artículos más representativos.

XI.A5 Receptores de estrógenos negativos

La ausencia de sobreexpresión de receptores de estrógenos (ER-) ha sido reportada por varios autores como significativamente asociada con el diagnóstico de metástasis cerebrales. En la práctica clínica se piensa que las pacientes con CaMa HER2 + y TN son los inmunofenotipos que mayormente están asociados con el diagnóstico de MeCe y por definición, estos dos inmunofenotipos tienen ER -, por lo que no es extraño que el presente estudio y la literatura establezcan una mayor posibilidad de metástasis cerebrales en aquellas pacientes ER- tal y como se presenta en la Tabla D6.

Tabla D6. Resumen de estudios publicados que han determinado una asociación significativa entre receptores de estrógeno negativo (ER-) con diagnóstico de metástasis cerebrales en pacientes con cáncer de mama.

Análisis	Medidas de asociación	Referencia
AS	ER - 8.1% (200/2481) ER + 4.2% (229/5457) P <0.01	Pestalozzi et al 2006(71)
AS	ER- 82% (40/49) vs ER + 18% (9/49) P = 0.001	Hicks et al 2006(50)
AM	ER - vs + OR 2.8 (IC95% 2.1 - 3.7) P <0.001	Tham et al 2006(73)
AS	SLMeCe ER+ 42 meses ER- 36 meses P = 0.021	Saip et al(93)
AM	TN OR 1.4 (1.01 - 1.95) P = 0.04 HER2+ ER ± OR 2.7 (IC95% 2.0 - 3.66) P <0.001	Graesslin et al 2010(75)
AS	ER- vs + OR 4.06 (IC95% 2.39 - 6.91) P NR	Minisini et al 2013(94)
AS	ER+ HR 0.29 (0.19 - 0.45) P <0.001	Sosinska-Mielcarek et al 2013(87)
AM	ER+ referencia ER- OR 5.027 (IC95% 1.268-19.933) P = 0.005	Xue et al 2013(85)
AM	TN sHR 2.08 (IC95% 1.09 - 4.00) P = 0.03 HER2+ ER± sHR 3.41 (IC95% 1.91-6.45) P <0.01	Genre et al 2017(46)
AM	TN HR 4.42 (2.86 - 6.85) P <0.001 HER2+ ER± HR 2.53 (1.57 - 4.07) P <0.001	Hung et al 2014(77)
AM	RH+/HER2- Ref RH+/HER2+ OR 1.41 (1.14 - 1.73) P = 0.001 RH-/HER2+ OR 2.09 (1.66 - 2.64) P <0.001 TN OR 2.19 (1.81 - 2.66) P <0.001	Martin et al 2017(88)
AS	Mediana SLMeCe (P = 0.01) ER- 26 meses ER+ 61 meses	Brogi et al 2011(78)
AM	ER- HR 5.1 (IC95% 1.8-14.6) P = 0.003	Rudat et al 2014(95)
AS	RH+/HER2+ vs no P 0.01	Shen et al 2015(80)
AM	ER+ HR 0.32 (IC95% 0.15 - 0.065) P = 0.002 Pacientes HER2+	Gabos et al 2006(82)
AS	ER- 55.3% RE+50.2% P = 0.004 AM negativa	Heitz et al 2011(76)
AS	ER- vs + P 0.029 AM negativa	Dayan et al 2012(81)
AS	ER- vs + P = 0.033 AM negativa	Chow et al 2015(83)
AM	ER- OR 5.027 (IC95% 1.268-19.933) P = 0.005	Xue et al 2013(85)
AM	HER2+ OR 1.916 (IC95% 1.381 - 2.659) P <0.001 TN OR 1.749 (1.291 - 2.368) P <0.001	Kim et al 2018(23)
AS	ER- vs + OR 2.75 (IC95% 2.16 - 3.50) P <0.0001	Base de entrenamiento
AM	ER- vs. + OR 1.88 (1.31 - 2.69) P = 0.001	Base de entrenamiento
AM	ER- vs. + OR 2.02 (IC95% 1.24 - 3.29) P = 0.004	Base de validación

AS = Análisis simple (univariado), AM = Análisis múltiple (multivariado), HR = Hazzard ratio, OR = Razón de momio, IC95% = Intervalo de confianza al 95%, ER = Receptores de estrógeno, RH = Receptores hormonales (incluye ER y PR)

XI.A6 Porcentaje de expresión de Ki-67

La proteína Ki-67 está presente en todas las células que proliferan y es un marcador de proliferación utilizado con frecuencia en varias neoplasias incluyendo el CaMa. La expresión de Ki-67 puede ser medida utilizando el anticuerpo monoclonal MIB-1. Clínicamente, una expresión elevada de Ki-67 en CaMa se ha asociado a un incremento en el riesgo de recurrencia y a una disminución en la supervivencia global.

El gen que codifica para Ki-67 se ha incluido en distintas firmas de proliferación, estas firmas han identificado patrones de expresión de varios genes asociados con el pronóstico de pacientes con CaMa y la necesidad de emplear tratamiento sistémico; por ejemplo, el gen que codifica para la proteína Ki-67 es uno de los genes seleccionados en la firma Oncotype DX®, esta firma genética se utiliza para predecir el riesgo de recurrencia en pacientes con cáncer de mama temprano y la necesidad de que requieran tratamiento sistémico(96).

El primer estudio en encontrar una asociación clínica entre la expresión de Ki-67 y el diagnóstico de MeCe fue publicado por Ishihara *et al.* en el 2013(54); en este estudio, mencionan que el nivel de corte aún no está establecido y varía entre diversos autores. En el presente estudio, utilizamos tres niveles de corte o nudos como lo recomiendan algunos autores incluyendo Chapman en la página 229 de su libro Medical Risk Prediction Models. CRC Press, Aunque los estratos no fueron significativamente asociados estadísticamente con MeCe en el análisis de regresión logística múltiple, el incluir esta variable con esos valores de corte mejoró el desempeño del modelo, por lo que se decidió incluir así la variable, tal y como ha sido recomendado por diversos autores incluyendo a David W. Hosmer en la tercera edición de su libro Applied logistic regression página 90. En la tabla D.7 se resumen los estudios más relevantes que han encontrado una asociación entre la expresión de Ki67 con el diagnóstico de MeCe.

Tabla D7. Resumen de los estudios que han encontrado una asociación entre la expresión de la proteína Ki-67 con el diagnóstico de MeCe

Análisis	Medida de asociación	Referencia
AS	Ki-67 >10% 16/18 P = 0.007	Shao 2011(79)
AM	Ki-67 ≥30% vs <30 HR 3.9 (IC95% 1.2-12.9) P = 0.026	Ishihara 2013(54)
AS	Ki-67 ≤14 vs >14 OR 0.77 (IC95% 0.70 – 0.84) P NR	Minisini 2013(94)
AS	Ki-67 ≥14% HR 2.99 (IC95% 1.92 – 4.67) P <0.001	Sosińska-Mielcarek 2013(87)
AS	Ki-67 ≥14 P <0.0001	Martínez-Aranda 2017(97)
AS	Ki-67 <10 Referencia Ki-67 10-19 OR 2.17 (IC95% 1.01 – 4.65) P = 0.046 Ki-67 20-40 OR 6.10 (IC95% 3.05 – 12.20) P <0.0001 Ki 67 >40 OR 5.93 (IC95% 2.93 – 12.05) P <0.0001	Base de entrenamiento
AM	Ki-67 <10 Referencia Ki-67 10-19 OR 1.90 (IC95% 0.87 – 4.15) P = 0.105 Ki-67 20-40 OR 3.66 (IC95% 1.75 – 7.66) P = 0.001 Ki-67 >40 OR 2.36 (IC95% 1.06 – 5.22) P = 0.034	Base de entrenamiento
AM	Ki-67 <10 Referencia Ki-67 10-19 OR 1.78 (IC95% 0.63 – 5.00) P = 0.270 Ki-67 20-40 OR 1.62 (IC95% 0.59 – 4.42) P = 0.345 Ki-67 >40 OR 1.18 (IC95% 0.41 – 3.44) P = 0.750	Base de validación

AM = Análisis múltiple, AS = Análisis simple, CMM = Cáncer de mama metastásico, EC = Etapa clínica, HR = Hazzard ratio, IC95% = Intervalo de confianza al 95% OR = Razon de Momios

XI.A7 SBR

El puntaje de Scarff-Bloom-Richardson (SBR) y el SBR-modificado(64) evalúan la formación de túbulos, características nucleares de pleomorfismo y el índice mitótico; este puntaje es uno de los más utilizados por patólogos en el mundo en sus reportes y es uno de los principales factores pronósticos en CaMa. El puntaje de estas tres características se gradúa del 1-3, obteniendo un mínimo de SBR de 3 y un máximo de 9. Es común que, con el resultado de este puntaje, se divida en tres grupos: SBR I, incluye resultados con puntaje de SBR de 3 a 5 puntos, SBR II incluye resultados con puntaje de 6 o 7 y Grado III cuando obtengan un puntaje de 8 o 9.

A mayor grado del SBR peor es el pronóstico, esto ha sido medido como periodo libre de enfermedad y supervivencia global(65). Existe evidencia que el SBR puede ayudar a predecir la respuesta del tratamiento citotóxico de la siguiente manera: a mayor puntaje del SBR, mejor respuesta al tratamiento citotóxico. El SBR se utiliza en la clínica para saber el comportamiento del tumor, ya que a mayor grado SBR será más agresivo.

Estudios han comprobado que un mayor grado histológico medido con SBR está asociado a una mayor recurrencia del tumor, menor supervivencia libre de enfermedad y menor supervivencia global en todos los subtipos de CaMa ya que se asocia a mayor proliferación e invasión celular(19,98).

Existen estudios que han reportado que el grado histológico (SBR) se asocia con el diagnóstico de MeCe. En la Tabla D8 se resumen los más importantes. El estudio de Saip et al(93) reportó que tras el análisis simple la SLMeCe fue distinto entre el grado histológico III (34 meses), SBR II (48 meses) que en el SBR I (67 meses) P = 0.002. A manera general, es consistente en los estudios previos, que el SBR está asociado directamente con MeCe, es decir, a mayor SBR mayor posibilidad de MeCe. Es importante aclarar, que similar a otras variables incluidas en el modelo, a pesar de no estar significativamente asociada a MeCe estadísticamente, el incluir esta variable al modelo predictor, mejoró su desempeño.

Tabla D8. Resumen de los estudios que han encontrado una asociación entre el puntaje SBR con el diagnóstico de MeCe

Análisis	Medida de asociación	Referencia
AS	Grado 2 o 3 HR 3.79 (IC95% 1.70 – 8.45) P 0.008	Gabos 2006(99)
AS	Grado I 0%, Grado 2 39% (9/23), III 61% (14/23) P = 0.002	Hicks 2006(50)
AS	Tiempo a MeCe 0.002	Saip 2009(93)
AM	Grado 3 vs 1,2 OR 1.54 (IC95% 1.10 – 2.13) P = 0.01	Graesslin 2010(75)
AS	Grado I 0/18 II 2/18, III 16/18 P <0.001	Shao 2011(79)
AS	Grado 1 1(3%), 2 6(17%), 3 29(81%) P <0.001	Arvold 2012(100)
AS	Grado3 vs 2/1 OR 1.97 (IC95%1.17 – 3.33) P NR	Minisini 2013(94)
AS	Grado 3 (49%) vs 1,2 (27%) P = 0.017	Sosińska-Mielcarek 2013(87)
AS	Grado3 vs 1 HR 5.89 (IC95% 1.45-23.93) P = 0.013 AM negativo	Hung 2014(101)
AM	Grado 3 vs 1,2 OR 6.83 (IC95% 1.04 – 22.91) P = 0.002	Chow 2015(83)
AM	Grado III vs 1,2 HR 1.76 (0.96 – 3.23) P = 0.06	Genre 2017(102)
AS	Grado III vs 1,2 HR 1.64 (IC95% 1.12 – 2.40) P = 0.010	Azim 2018(39)
AM	Grado I referencia, II OR 1.52 (IC95% 0.87-2.67) P = 0.138, III OR 1.54 (IC95%0.87-2.71) P = 0.135	Kim 2018
AS	Grado I Referencia Grado II OR 1.10 (IC95% 0.75 – 1.60) P = 0.625 Grado III OR 2.45 (IC95% 1.71 – 3.52) P <0.0001	Base de entrenamiento
AM	Grado I Referencia Grado II OR 0.47 (IC95% 0.29 – 0.76) P = 0.002 Grado III OR 0.70 (IC95% 0.43 – 1.13) P = 0.155	Base de entrenamiento
AM	Grado I Referencia Grado II OR 2.22 (IC95% 0.89 – 5.52) P = 0.084 Grado III OR 3.05 (IC95% 1.21 – 7.70) P = 0.018	Base de validación

AM = Análisis múltiple, AS = Análisis simple, CMM = Cáncer de mama metastásico, HR = Hazzard ratio, IC95% = Intervalo de confianza al 95% OR = Razon de Momios

XI.A8 Estado de la expresión del HER2

La sobreexpresión del receptor del factor humano de crecimiento epidérmico tipo 2 (HER-2, ErbB-2) define un subtipo agresivo de cáncer de mama caracterizado por proliferación celular rápida, incremento en la angiogénesis, apoptosis deficiente y un incremento de formación de metástasis. La amplificación de HER-2 y/o sobreexpresión ocurre en alrededor del 25% de los cánceres de mama invasores. La activación del receptor de la tirosina cinasa HER-2 activa múltiples vías de señalización que regulan el crecimiento celular normal y promueve la génesis tumoral vía proliferación celular, supervivencia, migración, diferenciación y angiogénesis(99).

Varios estudios han reportado que la sobreexpresión de HER-2 (HER2+) es un factor pronóstico desfavorable al estar asociado con tumores de alto grado, incremento en la proliferación celular, e involucro de adenopatías. Además, HER2+ se ha asociado con resistencia al uso de tamoxifeno porque el nivel de receptores hormonales en estos tumores es menor que en aquellos HER2 negativos, posiblemente con disminución en la sensibilidad a la terapéutica con antraciclinas(99).

Estos factores contribuyen a un incremento en el riesgo de recurrencia a distancia, una peor supervivencia global, siendo una de las probables causas de este peor pronóstico un incremento en la prevalencia de metástasis cerebrales. Varios estudios han asociado HER2+ a un incremento en el riesgo de MeCe. Los principales estudios se describen en la Tabla D.6. Incluso algunos autores han propuesto dar radioterapia profiláctica a cráneo en pacientes con CaMa HER2+ con o sin metástasis a hígado(93). Sin embargo, al ajustar el estado de expresión de HER2 por las variables que fueron incluidas en el modelo final no resultó ser estadísticamente significativa [OR 0.802 (IC95% 0.579 – 1.11) P = 0.184].

El agregar esta variable al modelo predictor no mejoró el rendimiento como se aprecia en el modelo por lo que se decidió no incluirla al modelo.

Tabla D9. Resumen de los estudios que han encontrado una asociación entre la expresión de HER2 con el diagnóstico de MeCe

Análisis	Medida de asociación	Referencia
AS	HER2+ 27/301 (9%) P = 0.0001 HR 4.23 (IC95% 1.84 – 9.74) P = 0.0007	Gabos et al 2006(99)
AM	HER2+ HR 3.55 (IC95% 1.45-8.72) P = 0.006	
AS	HER2+ 24%. (12/51) P = 0.001	Hicks et al 2006(50)
AM	HER2+ HR 3.4 (IC95% 3.1 – 10.9) P = 0.005	Heitz et al 2009(74)
AS	HER2+ AUC 0.76 (IC95% 0.60 – 0.93) P NR	Sanz-Pamplona et al 2011(103)
AS	HER2+ SLMeCe 34.5 meses, 3/55 (5%)	Arvold et al 2012(100)
AM	HER2+ P = 0.001	Dayan et al 2012(81)
AM	HER2+ OR 4.47 (IC95% 1.45 – 13.79) P NR	Minisini et al 2013(94)
AS	HER2+ HR 1.75 (IC95% 1.11 -2.76) P 0.016	Sosińska-Mielcarek et al 2013(87)
AM	HER2+ referencia HER2- OR 7.039 (IC95% 1.109-23.043) P = 0.005	Xue et al 2013(85)
AM	HER2+ vs LumA HR 2.53 (IC95% 1.57-4.07) P <0.001	Hung et al 2014(101)
AS	HER2+ RR 2-489 (IC95%1.07-5.78) P = 0.030	Wangchinda et al 2016(104)
AS	HER2+ HR 3.5 (95%CI 1.1-8.3) P = 0.001	Lim et al 2017(105)
AS	HER2+ 12/21(57%) P = 0.01	Martínez-Aranda et al 2017(97)
AM	HER2+ HR 3.41 (IC95% 1.91 – 6.45) P <0.01	Genre 2017(102)
AM	HER2+ HR 1.89 (IC95% 1.00 – 3.41) P = 0.049	Azim et al 2018(106)
AS	HER2+ OR 2.39 (IC95% 1.86 – 3.07) P <0.0001	Base de entrenamiento

AM = Análisis múltiple, AS = Análisis simple (Univariable), CMM = Cáncer de mama metastásico, EC = Etapa clínica, HR = Hazzard ratio, IC95% = Intervalo de confianza al 95%, NR = no reportada OR = Razon de Momios, RR = Riesgo relativo

XI.B DESEMPEÑO DEL MODELO PREDICTOR

Al evaluar un modelo predictor se debe considerar cuál es el objetivo de su uso, para el presente estudio se planeó generar una herramienta (modelo predictor) que permita identificar un grupo de alto riesgo de metástasis cerebrales para que en un futuro se pueda plantear una intervención de cribado, diagnóstico o terapéutica en este grupo de “alto riesgo de MeCe”; es decir, que tenga una buena “*capacidad de discriminación*” entre pacientes que serán diagnosticadas con metástasis cerebrales de aquellas que no, el AUC y estadístico C son considerados como los mejores medidores de la capacidad de discriminación.

La discriminación en un modelo predictor puede ser evaluada y se recomienda que sea evaluada con el estadístico C midiendo el área bajo la curva (AUC) generada entre la probabilidad obtenida por el modelo (probabilidad calculada) y la probabilidad observada (diagnóstico o no de MeCe). Como consenso general – ya que no existen guías exactas – se ha establecido un punto de corte para el $AUC \geq 0.7$ para determinar que existe una adecuada capacidad de discriminación del modelo predictor. En el presente estudio el AUC del modelo propuesto fue de 0.805 con un intervalo de confianza al 95% (IC95%) de 0.776 – 0.834 y con un estadístico C (significancia estadística) <0.0001 en la BaseE y los valores en la BaseV fueron: AUC 0.829 (IC95% 0.794 – 0.864) $P <0.0001$. Con estos resultados se corrobora que el modelo predictor propuesto tiene una buena capacidad de discriminación.

Al construir un modelo predictor, un error común es incluir solo variables que son “significativamente asociadas” con el desenlace (MeCe) en la regresión logística múltiple, si hubiéramos tomado esa decisión, el desempeño del modelo no sería el óptimo. Por ejemplo, si hubiéramos dejado fuera todas las variables que no fueron significativas tras el análisis múltiple como es el caso de SBR, el modelo hubiera tenido un AUC de 0.798 (IC95% 0.768 – 0.828) y un AIC de 1365 a diferencia de Modelo escogido que obtuvo un AUC de 0.805) y un AIC de 1354; ambos resultados en la BaseE.

Otra medida para evaluar el desempeño de un modelo predictor es la calibración. Se recomienda hacer esta medición en ambas bases (BaseE y BaseV) para confirmar la capacidad predictora del modelo (desempeño del modelo).

Calibración: Se conoce también como la confiabilidad del modelo para la población a la cual se va a emplear (población a estudiar = mujeres con cáncer de mama), se puede medir con el criterio de información de Akaike (AIC), se considera que a menor puntaje mejor calibración, aunque no existe tampoco un nivel de corte preestablecido. El AIC para el modelo propuesto fue de 1354 en la Base de entrenamiento y 682 en la base de validación.

Al compararlo la calibración con otros modelos; por ejemplo, el Neo-Bioscore que predice pronóstico en mujeres con cáncer de mama después de un tratamiento neoadyuvante cuyo AIC fue de 3333.38(107) el modelo que se propone pareciera tener una adecuada calibración.

Interpretación del Modelo

Se generó una calculadora en Excel para la medición de la posibilidad individual (de cada paciente) de MeCe con las variables comentadas cuya fórmula se generó con el resultado de la constante (β) general y de cada variable como se describe en la sección R.8 y así facilitar el uso en la práctica diaria.

Validación interna

Los procedimientos realizados para que el modelo predictor pueda medir lo que realmente queremos medir (validez interna) fueron:

- a) Dividir la cohorte en una BaseE y una BaseV con una división simple al generar número aleatorios 1-1-2, obteniendo de una BaseE con 3369 y una BaseV con 1684 pacientes con CaMa.
- b) El modelo predictor se creó en la BaseE y se confirmó el desempeño en la BaseV.
- c) Se realizó una validación cruzada por 10 (10-fold cross validation) del modelo en la BaseE, encontrando que en las 10 corridas del modelo se conservó una AUC ≥ 0.70 que es el corte recomendado.
- d) Se capacitó al personal que ayudó al llenado de la base de datos, se estandarizaron y definieron a priori las variables y se supervisó la captura de datos semanalmente. En caso de que existiera alguna duda, se acudió con expertos en el área de oncología médica y/o patología para resolver cada duda.
- e) Se consideró estudiar una cohorte con un gran número de tamaño de la muestra.
- f) Se calcularon los valores faltantes para excluir la menor cantidad de pacientes.

Validación externa

Con la finalidad de confirmar la generalización del modelo propuesto, la evidencia apunta a que se requiere validar en una cohorte externa el modelo y así completar el proceso de validación del modelo. En estos momentos, se está trabajando con dos instituciones (INCMNSZ e ITESM) para juntar el tamaño de la muestra requerido para medir el desempeño del modelo en una cohorte externa.

XII. CONCLUSIÓN

Un modelo predictor de metástasis cerebrales construido con variables clínicas (edad, tamaño tumoral, presencia de adenopatía axilar y etapa clínica anatómica) y variables patológicas (ER, Ki67 y SBR) logró obtener una adecuada capacidad discriminadora y calibración en una cohorte de 6,019 mujeres con cáncer de mama mexicanas atendidas en el INCAN.

XIII. Referencias bibliográficas

1. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015 Jan 6;162(1):W1-73.
2. Kattan MW, Hess KR, Amin MB, Lu Y, Moons KGM, Gershengwald JE, et al. American Joint Committee on Cancer acceptance criteria for inclusion of risk models for individualized prognosis in the practice of precision medicine. *CA Cancer J Clin*. 2016 Sep;66(5):370-4.
3. Bargalló Rocha, Juan Enrique, Shaw Dubin, Robin, Matus Santos, Juan Antonio, et al. Carcinoma invasor de la mama. In: Tratamiento del cáncer Oncología médica, quirúrgica y radioterapia. 2a ed. México: Manual Moderno; 2020. p. 856.
4. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000 Aug 17;406(6797):747-52.
5. ABRAMS HL, SPIRO R, GOLDSTEIN N. Metastases in carcinoma; analysis of 1000 autopsied cases. *Cancer*. 1950 Jan;3(1):74-85.
6. LESSE S, NETSKY MG. Metastasis of neoplasms to the central nervous system and meninges. *AMA Arch Neurol Psychiatry*. 1954 Aug;72(2):133-53.
7. ARONSON SM, GARCIA JH, ARONSON BE. METASTATIC NEOPLASMS OF THE BRAIN: THEIR FREQUENCY IN RELATION TO AGE. *Cancer*. 1964 May;17:558-63.
8. Viadana E, Cotter R, Pickren JW, Bross ID. An autopsy study of metastatic sites of breast cancer. *Cancer Res*. 1973 Jan;33(1):179-81.
9. Posner JB, Chernik NL. Intracranial metastases from systemic cancer. *Adv Neurol*. 1978;19:579-92.
10. Cifuentes N, Pickren JW. Metastases from carcinoma of mammary gland: an autopsy study. *J Surg Oncol*. 1979;11(3):193-205.
11. Cho SY, Choi HY. Causes of death and metastatic patterns in patients with mammary cancer. Ten-year autopsy study. *Am J Clin Pathol*. 1980 Feb;73(2):232-4.
12. Hagemester FBJ, Buzdar AU, Luna MA, Blumenschein GR. Causes of death in breast cancer: a clinicopathologic study. *Cancer*. 1980 Jul 1;46(1):162-7.
13. Amer MH. Chemotherapy and pattern of metastases in breast cancer patients. *J Surg Oncol*. 1982 Feb;19(2):101-5.
14. Tsukada Y, Fouad A, Pickren JW, Lane WW. Central nervous system metastasis from breast carcinoma. Autopsy study. *Cancer*. 1983 Dec 15;52(12):2349-54.
15. Lamovec J, Zidar A. Association of leptomeningeal carcinomatosis in carcinoma of the breast with infiltrating lobular carcinoma. An autopsy study. *Arch Pathol Lab Med*. 1991 May;115(5):507-10.
16. de la Monte SM, Hutchins GM, Moore GW. Influence of age on the metastatic behavior of breast carcinoma. *Hum Pathol*. 1988 May;19(5):529-34.
17. Clark GM, Sledge GWJ, Osborne CK, McGuire WL. Survival from first recurrence: relative importance of prognostic factors in 1,015 breast cancer patients. *J Clin Oncol Off J Am Soc Clin Oncol*. 1987 Jan;5(1):55-61.
18. Martin AM, Cagney DN, Catalano PJ, Warren LE, Bellon JR, Punglia RS, et al. Brain Metastases in Newly Diagnosed Breast Cancer: A Population-Based Study. *JAMA Oncol*. 2017 Aug 1;3(8):1069-77.
19. Largillier R, Ferrero JM, Doyen J, Barriere J, Namer M, Mari V, et al. Prognostic

- factors in 1,038 women with metastatic breast cancer. *Ann Oncol Off J Eur Soc Med Oncol*. 2008 Dec;19(12):2012–9.
20. Klar N, Rosenzweig M, Diergaard B, Brufsky A. Features Associated With Long-Term Survival in Patients With Metastatic Breast Cancer. *Clin Breast Cancer*. 2019 Aug;19(4):304–10.
21. Mohar-Betancourt A, Alvarado-Miranda A, Torres-Domínguez JA, Cabrera P, Lara Medina F, Villarreal-Gómez YS, et al. Factores pronósticos en pacientes con cáncer de mama y metástasis cerebral como primer sitio de recurrencia. *Salud Pública México Vol 60 Núm 2mar-Abr 2018* [Internet]. 2018; Available from: <http://www.saludpublica.mx/index.php/spm/article/view/9082>
22. Li S, Zhao J, Zhu L, Su F, Chen K. Development and validation of a nomogram predicting the overall survival of stage IV breast cancer patients. *Cancer Med*. 2017 Nov;6(11):2586–94.
23. Kim YJ, Kim JS, Kim IA. Molecular subtype predicts incidence and prognosis of brain metastasis from breast cancer in SEER database. *J Cancer Res Clin Oncol* [Internet]. 2018 Sep 1;144(9):1803–16. Available from: <https://doi.org/10.1007/s00432-018-2697-2>
24. Kallioniemi OP, Holli K, Visakorpi T, Koivula T, Helin HH, Isola JJ. Association of c-erbB-2 protein over-expression with high rate of cell proliferation, increased risk of visceral metastasis and poor long-term survival in breast cancer. *Int J Cancer*. 1991 Nov 11;49(5):650–5.
25. Aversa C, Rossi V, Geuna E, Martinello R, Milani A, Redana S, et al. Metastatic breast cancer subtypes and central nervous system metastases. *The Breast* [Internet]. 2014 Oct 1 [cited 2022 Apr 12];23(5):623–8. Available from: <https://doi.org/10.1016/j.breast.2014.06.009>
26. Arvold ND, Oh KS, Niemierko A, Taghian AG, Lin NU, Abi-Raad RF, et al. Brain metastases after breast-conserving therapy and systemic therapy: incidence and characteristics by biologic subtype. *Breast Cancer Res Treat*. 2012 Nov;136(1):153–60.
27. Slimane K, Andre F, Delaloge S, Dunant A, Perez A, Grenier J, et al. Risk factors for brain relapse in patients with metastatic breast cancer. *Ann Oncol Off J Eur Soc Med Oncol*. 2004 Nov;15(11):1640–4.
28. Kuksis M, Gao Y, Tran W, Hoey C, Kiss A, Komorowski AS, et al. The Incidence of Brain Metastases Among Patients with Metastatic Breast Cancer: A Systematic Review and Meta-Analysis. *Neuro-Oncol* [Internet]. 2020 Dec 23 [cited 2021 Jan 5];noaa285. Available from: <https://academic.oup.com/neuro-oncology/advance-article/doi/10.1093/neuonc/noaa285/6046194>
29. Crivellari D, Pagani O, Veronesi A, Lombardi D, Nolè F, Thürlimann B, et al. High incidence of central nervous system involvement in patients with metastatic or locally advanced breast cancer treated with epirubicin and docetaxel. *Ann Oncol Off J Eur Soc Med Oncol*. 2001 Mar;12(3):353–6.
30. Miller KD, Weathers T, Haney LG, Timmerman R, Dickler M, Shen J, et al. Occult central nervous system involvement in patients with metastatic breast cancer: prevalence, predictive factors and impact on overall survival. *Ann Oncol Off J Eur Soc Med Oncol*. 2003 Jul;14(7):1072–7.
31. Clayton AJ, Danson S, Jolly S, Ryder WDJ, Burt PA, Stewart AL, et al. Incidence of cerebral metastases in patients treated with trastuzumab for metastatic breast cancer. *Br J Cancer*. 2004 Aug 16;91(4):639–43.

32. Gonzalez-Angulo AM, Cristofanilli M, Strom EA, Buzdar AU, Kau SW, Broglio KR, et al. Central nervous system metastases in patients with high-risk breast carcinoma after multimodality treatment. *Cancer*. 2004 Oct 15;101(8):1760–6.
33. Lai R, Dang CT, Malkin MG, Abrey LE. The risk of central nervous system metastases after trastuzumab therapy in patients with breast carcinoma. *Cancer*. 2004 Aug 15;101(4):810–6.
34. Evans AJ, James JJ, Cornford EJ, Chan SY, Burrell HC, Pinder SE, et al. Brain metastases from breast cancer: identification of a high-risk group. *Clin Oncol R Coll Radiol G B*. 2004 Aug;16(5):345–9.
35. Altaha R, Crowell E, Hobbs G, Higa G, Abraham J. Increased risk of brain metastases in patients with HER-2/neu-positive breast carcinoma. *Cancer*. 2005 Feb 1;103(3):442–3.
36. Ryberg M, Nielsen D, Osterlind K, Andersen PK, Skovsgaard T, Dombernowsky P. Predictors of central nervous system metastasis in patients with metastatic breast cancer. A competing risk analysis of 579 patients treated with epirubicin-based chemotherapy. *Breast Cancer Res Treat*. 2005 Jun;91(3):217–25.
37. Geyer CE, Forster J, Lindquist D, Chan S, Romieu CG, Pienkowski T, et al. Lapatinib plus Capecitabine for HER2-Positive Advanced Breast Cancer. *N Engl J Med* [Internet]. 2006 Dec 28 [cited 2020 Jul 23];355(26):2733–43. Available from: <https://doi.org/10.1056/NEJMoa064320>
38. Brufsky AM, Mayer M, Rugo HS, Kaufman PA, Tan-Chiu E, Tripathy D, et al. Central nervous system metastases in patients with HER2-positive metastatic breast cancer: incidence, treatment, and survival in patients from registHER. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2011 Jul 15;17(14):4834–43.
39. Azim HA, Abdel-Malek R, Kassem L. Predicting Brain Metastasis in Breast Cancer Patients: Stage Versus Biology. *Clin Breast Cancer*. 2018 Apr;18(2):e187–95.
40. Mohar-Betancourt A, Alvarado-Miranda A, Torres-Domínguez JA, Cabrera P, Lara-Medina F, Villarreal-Gómez YS, et al. [Prognostic factors in patients with breast cancer and brain metastasis as the first site of recurrence]. *Salud Publica Mex*. 2018 Apr;60(2):141–50.
41. Jin J, Gao Y, Zhang J, Wang L, Wang B, Cao J, et al. Incidence, pattern and prognosis of brain metastases in patients with metastatic triple negative breast cancer. *BMC Cancer*. 2018 Apr 19;18(1):446.
42. Cacho-Díaz B, Salmerón-Moreno K, Alvarez-Alvarez A, Mendoza-Olivas LG, Alvarado-Miranda A, Villarreal-Garza C, et al. Identification of risk factors for central nervous system metastasis in patients with breast cancer with neurologic symptoms. *Cancer*. 2020 May 26;
43. Komorowski AS, Warner E, MacKay HJ, Sahgal A, Pritchard KI, Jerzak KJ. Incidence of Brain Metastases in Nonmetastatic and Metastatic Breast Cancer: Is There a Role for Screening? *Clin Breast Cancer*. 2020 Feb;20(1):e54–64.
44. Cacho-Díaz B, Spinola-Marono H, Arrieta VA, Granados-García M, Wegman-Ostrosky T, Mendoza-Olivas LG, et al. Diagnosis of brain metastases in breast cancer patients resulting from neurological symptoms. *Clin Neurol Neurosurg*. 2018 Oct;173:61–4.
45. Graesslin O, Abdulkarim BS, Coutant C, Huguet F, Gabos Z, Hsu L, et al. Nomogram to predict subsequent brain metastasis in patients with metastatic breast cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 2010 Apr 20;28(12):2032–7.

46. Genre L, Roche H, Varela L, Kanoun D, Ouali M, Filleron T, 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 Oxf Engl* 1990. 2017 Feb;72:200–9.
47. Canney P, Murray E, Dixon-Hughes J, Lewsley LA, Paul J. A Prospective Randomised Phase III Clinical Trial Testing the Role of Prophylactic Cranial Radiotherapy in Patients Treated with Trastuzumab for Metastatic Breast Cancer - Anglo Celtic VII. *Clin Oncol R Coll Radiol G B*. 2015 Aug;27(8):460–4.
48. Huang F, Alrefae M, Langleben A, Roberge D. Prophylactic Cranial Irradiation in Advanced Breast Cancer: A Case for Caution. *Int J Radiat Oncol [Internet]*. 2009 Mar 1;73(3):752–8. Available from: <https://www.sciencedirect.com/science/article/pii/S0360301608009310>
49. Niwińska A, Tacikowska M, Murawska M. The effect of early detection of occult brain metastases in HER2-positive breast cancer patients on survival and cause of death. *Int J Radiat Oncol Biol Phys*. 2010 Jul 15;77(4):1134–9.
50. Hicks DG, Short SM, Prescott NL, Tarr SM, Coleman KA, Yoder BJ, 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*. 2006 Sep;30(9):1097–104.
51. Fredholm H, Magnusson K, Lindström LS, Tobin NP, Lindman H, Bergh J, et al. Breast cancer in young women and prognosis: How important are proliferation markers? *Eur J Cancer Oxf Engl* 1990. 2017 Oct;84:278–89.
52. de Azambuja E, Cardoso F, de Castro GJ, Colozza M, Mano MS, Durbecq V, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. *Br J Cancer*. 2007 May 21;96(10):1504–13.
53. Smith I, Robertson J, Kilburn L, Wilcox M, Evans A, Holcombe C, et al. Long-term outcome and prognostic value of Ki67 after perioperative endocrine therapy in postmenopausal women with hormone-sensitive early breast cancer (POETIC): an open-label, multicentre, parallel-group, randomised, phase 3 trial. *Lancet Oncol*. 2020 Nov;21(11):1443–54.
54. Ishihara M, Mukai H, Nagai S, Onozawa M, Nihei K, Shimada T, 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 [Internet]*. 2013;84(3):135–40. Available from: <https://www.karger.com/DOI/10.1159/000345321>
55. Jalava P, Kuopio T, Juntti-Patinen L, Kotkansalo T, Kronqvist P, Collan Y. Ki67 immunohistochemistry: a valuable marker in prognostication but with a risk of misclassification: proliferation subgroups formed based on Ki67 immunoreactivity and standardized mitotic index. *Histopathology*. 2006 May;48(6):674–82.
56. Goldhirsch A, Ingle JN, Gelber RD, Coates AS, Thürlimann B, Senn HJ. Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009. *Ann Oncol Off J Eur Soc Med Oncol*. 2009 Aug;20(8):1319–29.
57. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. *J Clin Oncol Off J Am Soc Clin*

Oncol. 2011 Nov 10;29(32):4273–8.

58. Tökés AM, Szász AM, Geszti F, Lukács LV, Kenessey I, Turányi E, et al. Expression of proliferation markers Ki67, cyclin A, geminin and aurora-kinase A in primary breast carcinomas and corresponding distant metastases. *J Clin Pathol*. 2015 Apr;68(4):274–82.

59. Gerdes J, Li L, Schlueter C, Duchrow M, Wohlenberg C, Gerlach C, et al. Immunobiochemical and molecular biologic characterization of the cell proliferation-associated nuclear antigen that is defined by monoclonal antibody Ki-67. *Am J Pathol*. 1991 Apr;138(4):867–73.

60. Menon SS, Guruvayoorappan C, Sakthivel KM, Rasmi RR. Ki-67 protein as a tumour proliferation marker. *Clin Chim Acta Int J Clin Chem*. 2019 Apr;491:39–45.

61. Gerds TA, Kattan MW. I am going to make a prediction model What do I need to know? In: *Medical Risk Prediction Models: With ties to machine learning*. First. CRC Press; 2021. p. 18.

62. Langholz B, Thomas DC. Nested case-control and case-cohort methods of sampling from a cohort: a critical comparison. *Am J Epidemiol*. 1990 Jan;131(1):169–76.

63. van Smeden M, Moons KG, de Groot JA, Collins GS, Altman DG, Eijkemans MJ, et al. Sample size for binary logistic prediction models: Beyond events per variable criteria. *Stat Methods Med Res*. 2019 Aug;28(8):2455–74.

64. Elston CW, Ellis IO. 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*. 1991 Nov;19(5):403–10.

65. Amat S, Penault-Llorca F, Cure H, Le Bouedec G, Achard JL, Van Praagh I, et al. Scarff-Bloom-Richardson (SBR) grading: a pleiotropic marker of chemosensitivity in invasive ductal breast carcinomas treated by neoadjuvant chemotherapy. *Int J Oncol*. 2002 Apr;20(4):791–6.

66. Jansen ACM, van Aalst-Cohen ES, Hutten BA, Büller HR, Kastelein JJP, Prins MH. Guidelines were developed for data collection from medical records for use in retrospective analyses. *J Clin Epidemiol [Internet]*. 2005 Mar 1;58(3):269–74. Available from: <https://www.sciencedirect.com/science/article/pii/S0895435604002823>

67. Xu Y, Goodacre R. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *J Anal Test*. 2018;2(3):249–62.

68. Korjus K, Hebart MN, Vicente R. An Efficient Data Partitioning to Improve Classification Performance While Keeping Parameters Interpretable. *PloS One*. 2016;11(8):e0161788.

69. Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast cancer. *The Lancet [Internet]*. [cited 2021 Apr 6]; Available from: [https://doi.org/10.1016/S0140-6736\(20\)32381-3](https://doi.org/10.1016/S0140-6736(20)32381-3)

70. Pivot X, Manikhas A, Żurawski B, Chmielowska E, Karaszewska B, Allerton R, et al. CEREBEL (EGF111438): A Phase III, Randomized, Open-Label Study of Lapatinib Plus Capecitabine Versus Trastuzumab Plus Capecitabine in Patients With Human Epidermal Growth Factor Receptor 2-Positive Metastatic Breast Cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 2015 May 10;33(14):1564–73.

71. Pestalozzi BC, Zahrieh D, Price KN, Holmberg SB, Lindtner J, Collins J, 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 Off J Eur Soc Med Oncol*. 2006 Jun;17(6):935–44.
72. Barnholtz-Sloan JS, Sloan AE, Davis FG, Vigneau FD, Lain P, Sawaya RE. Incidence Proportions of Brain Metastases in Patients Diagnosed (1973 to 2001) in the Metropolitan Detroit Cancer Surveillance System. *J Clin Oncol*. 2004 Jul;22(14):2865–72.
73. Tham YL, Sexton K, Kramer R, Hilsenbeck S, Elledge R. Primary breast cancer phenotypes associated with propensity for central nervous system metastases. *Cancer*. 2006 Aug 15;107(4):696–704.
74. Heitz F, Harter P, Lueck HJ, Fissler-Eckhoff A, Lorenz-Salehi F, Scheil-Bertram S, et al. Triple-negative and HER2-overexpressing breast cancers exhibit an elevated risk and an earlier occurrence of cerebral metastases. *Eur J Cancer Oxf Engl 1990*. 2009 Nov;45(16):2792–8.
75. Graesslin O, Abdulkarim BS, Coutant C, Huguet F, Gabos Z, Hsu L, et al. Nomogram to predict subsequent brain metastasis in patients with metastatic breast cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 2010 Apr 20;28(12):2032–7.
76. Heitz F, Rochon J, Harter P, Lueck HJ, Fissler-Eckhoff A, Barinoff J, et al. Cerebral metastases in metastatic breast cancer: disease-specific risk factors and survival. *Ann Oncol [Internet]*. 2011 Jul 1;22(7):1571–81. Available from: <https://www.sciencedirect.com/science/article/pii/S0923753419384315>
77. Hung MH, Liu CY, Shiau CY, Hsu CY, Tsai YF, Wang YL, et al. Effect of Age and Biological Subtype on the Risk and Timing of Brain Metastasis in Breast Cancer Patients. *PLOS ONE*. 2014 Feb 24;9(2):e89389.
78. Brogi E, Murphy CG, Johnson ML, Conlin AK, Hsu M, Patil S, et al. Breast carcinoma with brain metastases: clinical analysis and immunoprofile on tissue microarrays. *Ann Oncol Off J Eur Soc Med Oncol*. 2011 Dec;22(12):2597–603.
79. Shao M min, Liu J, Vong JS, Niu Y, Germin B, Tang P, et al. A subset of breast cancer predisposes to brain metastasis. *Med Mol Morphol*. 2011 Mar;44(1):15–20.
80. Shen Q, Sahin AA, Hess KR, Suki D, Aldape KD, Sawaya R, et al. Breast Cancer With Brain Metastases: Clinicopathologic Features, Survival, and Paired Biomarker Analysis. *The Oncologist*. 2015 May;20(5):466–73.
81. Dayan A, Koca D, Akman T, Oztop I, Ellidokuz H, Yilmaz U. The factors that have an impact on the development of brain metastasis in the patients with breast cancer [Internet]. Vol. 8, *Journal of Cancer Research and Therapeutics*. Wolters Kluwer Medknow Publications; 2012. p. 542–8. Available from: <https://www.cancerjournal.net/article.asp?issn=0973-1482;year=2012;volume=8;issue=4;spage=542;epage=548;aulast=Dayan>
82. Gabos Z, Sinha R, Hanson J, Chauhan N, Hugh J, Mackey JR, 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 Off J Am Soc Clin Oncol*. 2006 Dec 20;24(36):5658–63.
83. Chow L, Suen D, Ma KK, Kwong A. Identifying risk factors for brain metastasis in breast cancer patients: Implication for a vigorous surveillance program. *Asian J Surg*. 2015 Oct;38(4):220–3.
84. Ishihara M, Mukai H, Nagai S, Onozawa M, Nihei K, Shimada T, 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*. 2013;84(3):135–40.

85. Xue J, Peng G, Yang JS, Ding Q, Cheng J. Predictive factors of brain metastasis in patients with breast cancer. *Med Oncol* [Internet]. 2013 Feb 12;30(1):337. Available from: <https://doi.org/10.1007/s12032-012-0337-2>

86. Minisini AM, Moroso S, Gerratana L, Giangreco M, Iacono D, Poletto E, et al. Risk factors and survival outcomes in patients with brain metastases from breast cancer. *Clin Exp Metastasis*. 2013 Dec;30(8):951–6.

87. Sosińska-Mielcarek K, Duchnowska R, Winczura P, Badzio A, Majewska H, Lakomy J, et al. Immunohistochemical prediction of brain metastases in patients with advanced breast cancer: The role of Rad51. *The Breast* [Internet]. 2013;22(6):1178–83. Available from:

<http://www.sciencedirect.com/science/article/pii/S0960977613002452>

88. Martin AM, Cagney DN, Catalano PJ, Warren LE, Bellon JR, Punglia RS, et al. Brain Metastases in Newly Diagnosed Breast Cancer: A Population-Based Study. *JAMA Oncol* [Internet]. 2017 Aug 1 [cited 2022 Apr 15];3(8):1069–77. Available from: <https://doi.org/10.1001/jamaoncol.2017.0001>

89. Shen Q, Sahin AA, Hess KR, Suki D, Aldape KD, Sawaya R, et al. Breast cancer with brain metastases: clinicopathologic features, survival, and paired biomarker analysis. *The oncologist*. 2015 May;20(5):466–73.

90. Martínez-Aranda A, Hernández V, Moreno F, Baixeras N, Cuadras D, Urruticoechea A, et al. Predictive and Prognostic Brain Metastases Assessment in Luminal Breast Cancer Patients: FN14 and GRP94 from Diagnosis to Prophylaxis. *Front Oncol*. 2017;7:283.

91. Sihto H, Lundin J, Lundin M, Lehtimäki T, Ristimäki A, Holli K, et al. Breast cancer biological subtypes and protein expression predict for the preferential distant metastasis sites: a nationwide cohort study. *Breast Cancer Res BCR*. 2011 Sep 13;13(5):R87.

92. Voduc KD, Nielsen TO, Perou CM, Harrell JC, Fan C, Kennecke H, et al. α B-crystallin expression in breast cancer is associated with brain metastasis. *Npj Breast Cancer* [Internet]. 2015 Oct 21;1(1):15014. Available from: <https://doi.org/10.1038/npjbcancer.2015.14>

93. Saip P, Cicin I, Eralp Y, Karagol H, Kucucuk S, Cosar Alas R, et al. Identification of patients who may benefit from the prophylactic cranial radiotherapy among breast cancer patients with brain metastasis. *J Neurooncol* [Internet]. 2009 Jun 1;93(2):243–51. Available from: <https://doi.org/10.1007/s11060-008-9769-0>

94. Minisini AM, Moroso S, Gerratana L, Giangreco M, Iacono D, Poletto E, et al. Risk factors and survival outcomes in patients with brain metastases from breast cancer. *Clin Exp Metastasis*. 2013 Dec;30(8):951–6.

95. Rudat V, El-Sweilmeen H, Brune-Erber I, Nour AA, Almasri N, Altuwaijri S, et al. Identification of breast cancer patients with a high risk of developing brain metastases: a single-institutional retrospective analysis. *BMC Cancer* [Internet]. 2014 Apr 24;14(1):289. Available from: <https://doi.org/10.1186/1471-2407-14-289>

96. de Azambuja E, Cardoso F, de Castro G, Colozza M, Mano MS, Durbecq V, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12 155 patients. *Br J Cancer* [Internet]. 2007 May 1;96(10):1504–13. Available from: <https://doi.org/10.1038/sj.bjc.6603756>

97. Martínez-Aranda A, Hernández V, Moreno F, Baixeras N, Cuadras D,

- Urruticoechea A, et al. Predictive and Prognostic Brain Metastases Assessment in Luminal Breast Cancer Patients: FN14 and GRP94 from Diagnosis to Prophylaxis. *Front Oncol* [Internet]. 2017;7. Available from: <https://www.frontiersin.org/article/10.3389/fonc.2017.00283>
98. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol Off J Am Soc Clin Oncol* [Internet]. 2009/02/09 ed. 2009 Mar 10;27(8):1160–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/19204204>
 99. Gabos Z, Sinha R, Hanson J, Chauhan N, Hugh J, Mackey JR, 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 Off J Am Soc Clin Oncol*. 2006 Dec 20;24(36):5658–63.
 100. Arvold ND, Oh KS, Niemierko A, Taghian AG, Lin NU, Abi-Raad RF, et al. Brain metastases after breast-conserving therapy and systemic therapy: incidence and characteristics by biologic subtype. *Breast Cancer Res Treat*. 2012 Nov;136(1):153–60.
 101. Hung MH, Liu CY, Shiau CY, Hsu CY, Tsai YF, Wang YL, et al. Effect of age and biological subtype on the risk and timing of brain metastasis in breast cancer patients. *PLoS One*. 2014;9(2):e89389.
 102. Genre L, Roché H, Varela L, Kanoun D, Ouali M, Filleron T, 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 Oxf Engl 1990*. 2017 Feb;72:200–9.
 103. Sanz-Pamplona R, Aragüés R, Driouch K, Martín B, Oliva B, Gil M, et al. Expression of endoplasmic reticulum stress proteins is a candidate marker of brain metastasis in both ErbB-2+ and ErbB-2- primary breast tumors. *Am J Pathol*. 2011 Aug;179(2):564–79.
 104. Wangchinda P, Ithimakin S. Factors that predict recurrence later than 5 years after initial treatment in operable breast cancer. *World J Surg Oncol*. 2016 Aug 24;14(1):223.
 105. Lim YJ, Lee SW, Choi N, Kwon J, Eom KY, Kang E, et al. Failure patterns according to molecular subtype in patients with invasive breast cancer following postoperative adjuvant radiotherapy: long-term outcomes in contemporary clinical practice. *Breast Cancer Res Treat*. 2017 Jun;163(3):555–63.
 106. Azim HA, Abdel-Malek R, Kassem L. Predicting Brain Metastasis in Breast Cancer Patients: Stage Versus Biology. *Clin Breast Cancer* [Internet]. 2018 Apr 1 [cited 2022 Apr 12];18(2):e187–95. Available from: <https://doi.org/10.1016/j.clbc.2017.08.004>
 107. Mittendorf EA, Vila J, Tucker SL, Chavez-MacGregor M, Smith BD, Symmans WF, et al. The Neo-Bioscore Update for Staging Breast Cancer Treated With Neoadjuvant Chemotherapy: Incorporation of Prognostic Biologic Factors Into Staging After Treatment. *JAMA Oncol*. 2016 Jul 1;2(7):929–36.

Otras Referencias para la generación de los Modelos predictores:

- A) Graf E, Schmoor C, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine* 1999; 18: 2529-2545
- B) Kattan MW, Gerds TA. The index of prediction accuracy: An intuitive measure useful for evaluating risk prediction models. *Diagnostic and Prognostic Research* 2018; 2(1): 7.
- C) Korn EL, Simon R. Measures of explained variation for survival data. *Statistics in Medicine* 1990; 9(5): 487-503.

- D) Tjur T. Coefficients of determination in logistic regression models – a new proposal: The coefficient of discrimination. *The American Statistician* 1996; 58(1): 267-288.
- E) Gerds TA, Kattan MW. *Medical Risk Prediction Models: With ties to machine learning*. 1st Ed. CRC Press. 2021. Pp 290.
- F) Gerds TA, Kattan MW. *Medical Risk Prediction Models: With ties to machine learning*. 1st Ed. CRC Press. 2021. Pp 187